It's a Price! It's a Value!

Economic Measurement Across Environments

Chae-Yeon Xon
Yonsei University

September 2026

Abstract

A retained price records one valuation of an object. Revaluation asks which further fields preserve its value in other environments. I study this choice through the target image of a source-record fiber. Stable mixed-derivative kernels give local common interaction coordinates, with a reference valuation carrying level information. In bilinear systems, the residual target map gives exact completion counts, an affine-prediction loss frontier, and the value of an additional field. Feasible environment and composition menus turn these benchmarks into record choices. A food-basket exercise compares those choices on a fixed accounting array; two composition indices complete four source valuations on its six-dimensional centered support. Treasury cash flows illustrate how calibration-optimal records perform in later environments. The analysis distinguishes common-state dimension, target-specific recovery, and evaluation of a selected rule. Fixed-rule intervals and loading-reselection ranges retain their respective uncertainty interpretations.

JEL Codes: C18, C43, C51, D11, D61, E01, G12

Keywords: economic measurement, representation, comparability, transportability, information sufficiency, dimension reduction, measurement design

Yonsei University. Email: eunixon@yonsei.ac.kr. This draft is circulated for discussion; comments and suggestions are welcome.

1. Introduction

Which records should an institution retain when the same object will be valued in different environments? A source expenditure records a basket at one price vector. Future revaluation may depend on composition that this number leaves open. For two quantities, the source expenditure $q_1 + q_2$ and the target expenditure $2q_1 + q_2$ distinguish different baskets. A composition contrast $q_1 - q_2$, or a suitable second price record, supplies the missing direction. The institution's choice is which scalar fields to carry forward for its intended valuation tasks.[1]

For object x, let $Y_O(x)$ collect the retained source valuations and $Y_T(x)$ the target valuations. The target values compatible with a record y form

$$\mathcal{R}_{O\to T}(y) = \{Y_T(x) : Y_O(x) = y\}. \tag{1}$$

A singleton gives exact recovery.[2] Variation within this set determines what an additional record must resolve. The source record fixes what is already known; the target specifies which remaining distinctions matter.

The paper connects that recovery question to a common representation of value. Theorem 1 states that exact recovery is equivalent to constancy on every source fiber and, under constant ranks, identifies the minimum number of unrestricted smooth fields needed locally for a selected target. Theorem 2 gives a second characterization: a family of environments has the stated local relational representation if and only if its bilateral mixed kernels have stable left and right null spaces of common rank. The resulting state carries the interaction coordinates together with a reference valuation.[3] Thus common-state dimension bounds the information available to record, while the target image of the source fiber determines what a particular use still requires. A rotating-kernel example and an additive-level example make these distinctions explicit.

For exact bilinear valuations, Theorem 3 turns the same residual target map into a sharp record frontier. Its rank is the exact number of completing linear fields, and the stated construction attains that count. Exposure moments and a target loss metric attach values to smaller records: the eigenvalue tail is the minimum affine-prediction loss at each field budget. A rank-one covariance update values a candidate record conditional on those already held. The economic recording problem specifies which loadings are feasible and what their preservation costs.

[1] A field is a real-valued function of inputs available when the record is created. The analysis distinguishes unrestricted smooth fields from the observed price and composition menus available in the applications.

[2] This is a global statement on the declared source fiber. The constant-rank result below gives the corresponding local coordinate count.

[3] Changing the reference environment reparameterizes the level coordinate; the interaction distribution remains the object disciplined by the mixed-kernel condition.

An environment menu permits observed price or discount loadings; a composition menu permits fields formed from the object's available inputs. The unrestricted linear frontier provides a benchmark for these choices. Greedy addition evaluates each next field, and exchange search revisits the chosen set. A complete subset optimum, a search rule, and its later performance answer different parts of the design question. I keep their calibration and evaluation samples explicit.[4] For a complete bilinear evaluation array, a fixed-rule loss identity also separates residual exposure variation from a shift in mean composition, both weighted by the later target environments.

The F-MAP exercise holds a food-basket accounting array fixed and changes the record delivered to an affine decoder. At four fields, greedy selection lowers held-out MSE by 71.7 percent relative to dispersed retention; exchange refinement gives 77.7 percent. For scale, the exchange rule's RMSE is \$0.1364 per normalized 100-kilogram basket.[5] Four independent source valuations leave two directions in the six-dimensional centered basket support. Two calibrated composition indices complete them to numerical precision. The comparison evaluates a constructed recording rule at the declared aggregation level. Its fixed-rule intervals and loading-reselection ranges refer to their stated uncertainty experiments.

Treasury cash flows provide a second setting with an exact bilinear valuation map. A residual certificate is optimal for the calibration moments and discount environments, while a duration record gives lower loss in the later sample. This comparison makes the population attached to the design criterion consequential. The frontier values a record for a specified task; subsequent evaluation asks how that choice travels.

Related literature. The joint design of models and measurement tools motivates the recording question (Almås et al. 2024). Measurement systems recover latent statistical objects from imperfect observations (Schennach 2022). Here the measurement map is declared, and the recovery object is a vector of valuations of the same object. Functional decomposition and aggregation supply the common-state lineage (Leontief 1947a,b; Nataf 1948; Green 1964; Goldman and Uzawa 1964; Gorman 1968, 1959); exact index numbers and Divisia path independence connect aggregation to valuation across environments (Diewert 1976; Gorman 1996). The local construction uses constant rank and common first integrals (Lee 2013; Chiappori and Ekeland 1999). Its bilateral mixed-kernel condition, reference field, and use in target-relative record counts are stated separately so that their relation to the classical results can be assessed.

[4] The complete subset optimum is an oracle comparison on a fixed menu. Greedy and exchange rules are implementable searches whose later loss is evaluated after their choices are frozen.

[5] The corresponding calibration expenditure benchmark is \$330.145 for the same normalized basket. The ratio supplies a scale; welfare conversion requires an additional valuation.

Sufficiency and dimension reduction describe information relevant to a response (Blackwell 1953; Lee et al. 2013; Hoderlein and Lewbel 2012). Active-subspace methods use derivatives to guide approximation (Constantine et al. 2014; Zahm et al. 2020). Goal-oriented experimental design values observations by the uncertainty remaining in a selected target (Attia et al. 2018; Zhong et al. 2026; Chen et al. 2025). In the Gaussian case, the average-loss criterion here is the zero-noise limit of the corresponding goal-oriented A-optimal criterion; Appendix C.3 gives the support-based derivation. Ky Fan and reduced-rank prediction supply the unrestricted spectral benchmark (Fan 1949; Tso 1981), and optimal-design and exchange-search methods supply the acquisition tools (Elfving 1952; Kiefer and Wolfowitz 1960; Wu et al. 2023; Lau and Zhou 2022). Capacity here counts permitted smooth or linear scalar fields; information-bottleneck methods instead specify an information-rate budget (Tishby et al. 1999). The paper brings these ingredients to a single recording problem: a common valuation state, the target variation left by the source record, and the loss achievable with the institution's permitted fields.

2. Retained records and revaluation

2.1. Objects, environments, and retained records

Let $\mathcal{X}$ be an economic-object space, $\mathcal{Z}$ an evaluation-environment space, and $m : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ a declared measurement map. Expenditure $m(q, p) = p'q$ and present value $m(c, z) = d(z)'c$ evaluate a fixed quantity vector or dated cash-flow vector in different price or discount environments. Units, currency conversions, and base-period normalizations are fixed before environments are compared.[6] Online appendix A states the corresponding normalization on equivalence classes.

DEFINITION 1 (Measurement profile). *The profile of x is $\mathfrak{M}(x) = [z \mapsto m(x, z)]$. Two objects are measurement equivalent, written $x \sim_M x'$, when their profiles agree. The profile image is $\mathcal{M} = \mathfrak{M}(\mathcal{X}/\sim_M)$.*

An institution retains a record from source environments $O = \{z_1, \dots, z_K\}$ and uses that record to value the same object at targets $T = \{z_1^\star, \dots, z_J^\star\}$. Define

$$(2) \qquad Y_O(x) = (m(x, z_1), \dots, m(x, z_K))', \qquad Y_T(x) = (m(x, z_1^\star), \dots, m(x, z_J^\star))'.$$

The capacity K counts retained scalar valuations at the stated numerical precision. A policy chooses their environments from a feasible menu. The target family and a loss on its valuation coordinates specify the use of the record. Here *transport* means recovery of

[6] For exact index-number representations and aggregation under additional demand restrictions, see Diewert (1976); Gorman (1968, 1996). The present object is the information required to revalue a fixed declared object across environments.

the same object's value across evaluation environments; its mathematical operation is factorization of evaluation maps.

DEFINITION 2 (Residual transport set). *For an attained source record $y \in Y_O(\mathcal{X})$, the residual transport set is*

$$\mathcal{R}_{O\to T}(y) = Y_T(Y_O^{-1}(y)) = \{Y_T(x) : Y_O(x) = y\}. \tag{3}$$

The source fiber $Y_O^{-1}(y)$ contains the objects compatible with the record, and its image under Y_T contains their compatible target valuations. The retained record can therefore determine the target vector while several objects remain compatible with it. Sufficiency is target-specific: a record that recovers one target family may leave positive residual variation for another.

2.2. A composition record for revaluation

Consider two goods, source prices $p_O = (1, 1)'$, and target prices $p_T = (2, 1)'$. Baskets $q^A = (3/2, 1/2)'$ and $q^B = (1/2, 3/2)'$ both cost two at the source; they cost 7/2 and 5/2 at the target. Hence $\mathcal{R}_{O\to T}(2)$ contains both target values. The scalar composition record $C(q) = q_1 - q_2$ completes the comparison:

$$p_T' q = \frac{3}{2} p_O' q + \frac{1}{2} C(q). \tag{4}$$

An environment with prices $(2, 1)'$ and the object-side certificate C supply the same missing valuation direction through different permitted measurements. The choice between them depends on the institution's measurement menu and available object-side inputs. This calculation holds the basket fixed. A cost-of-living or welfare comparison would additionally specify preferences and the compensating adjustment to the basket. That specification would define a different target map and, potentially, a different completing record.

2.3. Target-specific revaluation

Consider three-good expenditure on an open set of positive quantity vectors at all positive price vectors. Its profile has dimension three. Retain expenditures at $p_1 = (1, 1, 1)'$ and $p_2 = (2, 1, 1)'$. The target $p_T = (3, 2, 2)' = p_1 + p_2$ is recovered as $p_T' q = p_1' q + p_2' q$. The source fiber permits offsetting changes in the second and third quantities along $(0, 1, -1)'$ while both retained expenditures stay fixed. Thus a three-coordinate profile admits exact target revaluation from two source fields. The relevant sufficiency condition is constancy of the target on that fiber.

2.4. Timing and the evaluation oracle

At the recording date, a source policy selects O using a calibration sample and a declared target specification. The institution retains $Y_O(x)$ and, when permitted, a certificate $C(x)$ formed from a declared pre-target input $P(x)$. At the revaluation date, the decoder receives the retained record and the target environment. Reporting, survey, and archival systems motivate the capacity constraint: future valuation uses the records that the system carries forward.

An operational record contains the source valuations and one scalar field for each certificate. A shared dictionary supplies the source environments, certificate weights, units, and decoder calibration. Quantity projections become composition-index fields; cash-flow projections become retained exposure summaries. Capacity counts scalar fields per object, while the recorded population shares the dictionary. The measurement choice concerns the information carried by each additional field. The dimension bounds concern smooth or linear records in the specified class. An explicit bit budget, rounding rule, or measurement-noise model would add a precision constraint to that menu.

The empirical laboratory observes quantities and prices to construct the accounting value for every held-out pair. These primitive data serve as an evaluation oracle. At the recording date, permitted object inputs can be used to form the selected scalar fields. The later decoder uses the retained source values, those fields, and the shared calibration dictionary. The laboratory therefore evaluates the consequences of a reporting or retention rule. Its intervention is the information passed to the future decoder. Monetary collection costs and behavioral responses are additional application primitives.

The F-MAP application places this timing in a data-publication chain. Commercial scanner inputs enter an upstream construction process, ERS publishes aggregate dollar, gram, unit-value, and price-index fields, and a downstream user carries selected economic fields into later revaluation. The empirical policy treats published area–month environments as a feasible counterfactual reporting menu. USDA operating costs and current field-retention choices lie outside the reported effect. This separation places the institutional primitive in the declared reporting capacity and leaves its monetary cost available for an application-specific objective.

The residual set $\mathcal{R}_{O\to T}$ is used at three levels. If it is a singleton, recovery is exact. Under constant ranks, the dimension of its local target variation counts the additional smooth information required for recovery. A declared object set or exposure distribution, together with a target loss, turns the same variation into a valuation-loss frontier. Adding a source environment or certificate contracts this residual information.

2.5. What the retained record identifies

For the source and target maps in (2), define the *target-relative residual dimension* and its tangent quotient by

$$d_{O\to T}(x) = \operatorname{rank}\Big(DY_T(x)|_{\ker DY_O(x)}\Big), \tag{5}$$

$$\mathcal{Q}_{O\to T}(x) = \ker DY_O(x) \big/ \left(\ker DY_O(x) \cap \ker DY_T(x)\right). \tag{6}$$

The rank counts source-preserving object directions that change target valuations. The term *transport deficiency* refers to this differential dimension. Le Cam deficiency is a separate experiment-approximation distance defined through Markov kernels (Le Cam 1964).

THEOREM 1 (Exact recovery and minimum local information). *Globally, $Y_T = h \circ Y_O$ for a map on attained records exactly when every residual transport set is a singleton, equivalently,*

$$Y_O(x) = Y_O(x') \implies Y_T(x) = Y_T(x'). \tag{7}$$

Suppose Y_O and (Y_O, Y_T) have constant ranks near x. Then

$$\dim \mathcal{Q}_{O\to T}(x) = d_{O\to T}(x) = \operatorname{rank} D(Y_O, Y_T)(x) - \operatorname{rank} DY_O(x). \tag{8}$$

This is the dimension of the local branch of $\mathcal{R}_{O\to T}(Y_O(x))$ through $Y_T(x)$. Exact local smooth transport holds precisely when $d_{O\to T} = 0$. A smooth certificate $C(x) \in \mathbb{R}^k$ and smooth decoder satisfying

$$Y_T(x) = h(Y_O(x), C(x)) \tag{9}$$

require $k \geq d_{O\to T}$. A local certificate of dimension $d_{O\to T}$ exists in the unrestricted smooth object-side class.

The global equivalence is deterministic sufficiency. The local statement applies the constant-rank theorem to that same source fiber. For a target metric $W_T \succ 0$, the form

$$g_{O,T,x}([\xi],[\eta]) = DY_T(x)[\xi]' W_T DY_T(x)[\eta] \tag{10}$$

is well defined and positive definite on $\mathcal{Q}_{O\to T}(x)$. Dimension counts the missing coordinates, while this form measures the target loss attached to them.

PROOF OF THEOREM 1. If Y_T is constant on source fibers, define $h(y) = Y_T(x)$ using any x with $Y_O(x) = y$. Constancy makes the definition well defined. Conversely, a factorization $Y_T = h \circ Y_O$ makes each image of a source fiber a singleton.

Under constant ranks, a local source fiber is a submanifold with tangent space $\ker DY_O$. Restricting Y_T to this fiber has differential $DY_T|_{\ker DY_O}$ of rank $\operatorname{rank} D(Y_O, Y_T) - \operatorname{rank} DY_O$. The constant-rank theorem gives a local image branch of this dimension. The first isomorphism theorem identifies its tangent space with $\mathcal{Q}_{O\to T}$. In source coordinates (s, w), zero rank means $D_w Y_T = 0$ throughout a connected local plaque; hence $Y_T = h(s)$ smoothly. Formula (10) is independent of the representative and positive definite because its kernel has been quotiented out.

Under $Y_T = h(Y_O, C)$, every $\xi \in \ker DY_O$ satisfies $DY_T[\xi] = D_C h\, DC[\xi]$. The rank bound is

$$d_{O\to T} \le \operatorname{rank}(DC|_{\ker DY_O}) \le k.$$

For attainment, choose $d_{O\to T}$ fixed linear combinations of the target components whose derivatives supplement DY_O to the joint rank at the point. After shrinking the neighborhood, (Y_O, C) and (Y_O, Y_T) have the same differential kernel. The local fiber criterion then gives a smooth decoder. □

3. Common measurement states

A common state describes all valuations generated by an object. Its dimension can exceed the information needed for a particular source–target comparison. The scalar attained-range representation gives the corresponding ranking implication. The main result concerns the joint variation of objects and environments.

3.1. Stable interaction kernels

A measurement state $s(x)$ generates the profile when $m(x, z) = g(s(x), z)$ for every environment. When the profile image is a smooth manifold, write $r_M = \dim \mathcal{M}$. The chain rule gives a lower bound $k \ge r_M$ for a smooth k-coordinate state; a chart of an embedded constant-rank profile image attains it locally under the stated function-space and evaluation assumptions.

Fix a reference pair (x_0, z_0) and remove the additive object and environment terms:

$$\Delta_0 m(x, z) = m(x, z) - m(x, z_0) - m(x_0, z) + m(x_0, z_0). \tag{11}$$

Its mixed derivative is $J_M(x, z) = D_x D_z m(x, z)$. A common relational quotient with dimension equal to the mixed rank requires the annihilated directions to persist as the opposing argument varies.

Economically, an object direction in the left kernel preserves the marginal response of valuation to every environmental change. Kernel stability requires the same response-irrelevant directions at each environment. The right-kernel condition imposes the corresponding requirement on environmental directions across objects. These restrictions concern the interaction component of value. An object direction can preserve that interaction and still change the reference valuation; Corollary 1 accounts for this level information. The map m and its units are fixed throughout the calculation.

ASSUMPTION 1 (Smooth interaction and stable kernels). *The measurement map is C^3 on a product of C^3 manifold neighborhoods $N_X \times N_Z$. Its left kernel $K_X(x) = \ker_L J_M(x,z)$ is independent of z, and its right kernel $K_Z(z) = \ker_R J_M(x,z)$ is independent of x. These kernels are smooth distributions of constant codimension r.*

THEOREM 2 (Relational representation characterization). *Under assumption 1, K_X and K_Z are involutive. Locally there are r-coordinate submersions $u(x), v(z)$ with $\ker Du = K_X$, $\ker Dv = K_Z$, and*

$$m(x,z) = a(x) + b(z) + F(u(x), v(z)), \tag{12}$$

$$\Delta_0 m(x,z) = F_0(u(x), v(z)), \qquad F_0(u, v_0) = F_0(u_0, v) = 0. \tag{13}$$

The induced mixed derivative $D_u D_v F_0$ has rank r. Conversely, such a factorization through submersions with nonsingular induced mixed derivative has the stated stable kernels.

PROOF OF THEOREM 2. Let X_1, X_2 be smooth fields in K_X. For fixed z and $\eta \in T_z\mathcal{Z}$, set $f_{z,\eta}(x) = D_z m(x,z)[\eta]$. Kernel stability gives $X_j f_{z,\eta} = 0$ throughout the object neighborhood. Therefore $[X_1, X_2] f_{z,\eta} = 0$, so the bracket also belongs to K_X. Interchanging object and environment proves involutivity of K_Z. Smooth constant codimension permits application of Frobenius, yielding local submersions u, v with kernels K_X, K_Z.

For $\xi \in K_X$, the function $D_\xi m(x,z)$ has zero derivative in every environment direction. On a connected local environment plaque it is constant, so $D_\xi[m(x,z) - m(x,z_0)] = 0$. The symmetric argument shows that the cross-difference annihilates K_Z as well. It is constant on products of local plaques and factors as $\Delta_0 m = F_0(u,v)$. The choices $a(x) = m(x,z_0)$ and $b(z) = m(x_0,z) - m(x_0,z_0)$ give the asserted full representation. Differentiation gives

$$J_M = Du(x)' D_u D_v F_0(u,v) Dv(z). \tag{14}$$

Since Du, Dv, J_M all have rank r, the induced mixed derivative is nonsingular. Conversely, the displayed factorization with submersions and a nonsingular middle matrix makes the left and right kernels equal to $\ker Du$ and $\ker Dv$. They are independent of the opposing argument. □

Direct constant-rank construction. In local coordinates choose $\eta_1, \dots, \eta_r \in T_{z_0}\mathcal{Z}$ so that the covectors $J_M(x_0, z_0)[\,\cdot\,, \eta_i]$ are independent. Define

$$u_i(x) = D_z m(x, z_0)[\eta_i], \qquad i = 1, \dots, r. \tag{15}$$

Independence persists on a smaller object neighborhood. Thus u is a C^2 submersion of rank r. Its differential annihilates $K_X(x)$, and equal dimensions give $\ker Du(x) = K_X(x)$. Choose $\xi_1, \dots, \xi_r \in T_{x_0}\mathcal{X}$ symmetrically and set $v_j(z) = D_x m(x_0, z)[\xi_j]$. Then $\ker Dv(z) = K_Z(z)$. The constant-rank theorem gives local sections and connected plaques for these maps; the preceding cross-difference argument defines a C^2 function F_0. For $r = 0$, the quotient consists of a point and the interaction vanishes locally. This construction makes the local first-integral argument explicit.

The restrictions have distinct roles in this construction. Smoothness makes marginal environment responses differentiable functions of the object. Kernel stability gives them common first integrals, and constant codimension permits a local submersion. Nonsingularity of the induced mixed derivative fixes the quotient dimension. Global Euclidean coordinates additionally require descent and chart conditions.

COROLLARY 1 (Relational and full profile dimension). *Under the stated profile-manifold regularity, the relational profile has differential rank r. The full profile is locally generated by*

$$s_{z_0}(x) = (m(x, z_0), u(x)), \qquad r \le r_M \le r + 1. \tag{16}$$

At regular points $r_M = r$ exactly when $m(\cdot, z_0)$ locally factors through u.

For the map $m(x, z) = x_1 z + x_2$ at reference environment $z_0 = 0$, the interaction coordinate is x_1 and the reference value is x_2. The mixed rank is one; the full profile carries both coordinates. An object direction can leave every marginal environment response unchanged and still shift the valuation level. This is the role of the reference field in (16).

The full state expresses both source and target maps in the quotient coordinates. Retaining the reference value $m(x, z_0)$ fixes the level coordinate. Residual revaluation then depends on the interaction coordinates left unresolved by the remaining source records. The information frontier below measures precisely this unresolved part.

3.2. Rotation across environments

PROPOSITION 1 (Rotating-kernel example). *For*

$$m(x,z) = x_1 z + x_2 z^2, \qquad J_M(x,z) = \begin{pmatrix} 1 \\ 2z \end{pmatrix}, \tag{17}$$

the pointwise mixed rank is one, while the relational profile has differential rank two. The left kernel $\{(\xi_1, \xi_2) : \xi_1 + 2z\xi_2 = 0\}$ *varies with* z*. On a product neighborhood with a nondegenerate environment interval, a representation through scalar* C^2 *submersions* u, v *and a nonzero induced mixed derivative would imply a fixed left kernel, which contradicts the displayed expression.*

One environment activates one object direction; the collection of environments activates both coordinates. The common state (x_1, x_2) is two-dimensional. The example separates pointwise mixed rank from the dimension of a shared relational state. Integrated derivative Gram matrices also register both directions under an environment distribution with interval support, as in active-subspace analysis (Constantine et al. 2014; Zahm et al. 2020). Block projector comparisons assess a proposed common approximation through the Gram characterization and its inference procedure.

Locally, the quotient admits the expansion $F_0(u,v) = \widetilde{u}' A_0 \widetilde{v} + o(\|\widetilde{u}\| \|\widetilde{v}\|)$, where $A_0 = D_u D_v F_0(u_0, v_0)$ is nonsingular. Expenditure and fixed-income valuation already have exact bilinear primitives. Exact bilinearity yields the projection formulas below, while constant-rank smoothness yields the local information statement.

4. Residual information and record choice

4.1. From the common state to a target-specific record

Sufficiency is defined directly on the source fiber. The common-state result provides a reduced coordinate system in which the same test can be computed.

The representation in Theorem 2 gives $m(x,z) = \widetilde{m}(s_{z_0}(x), z)$, with $\widetilde{m}((\alpha, u), z) = \alpha + c(z) + F_0(u, v(z))$ and $c(z) = m(x_0, z) - m(x_0, z_0)$. At a regular local state image $S = s_{z_0}(N_X)$,

$$d_{O\to T}(x) = \operatorname{rank}\Big(D\widetilde{Y}_T|_{\ker D\widetilde{Y}_O} \Big). \tag{18}$$

The derivatives on the right act on $T_s S$. Equation (18) connects the common-state theorem to the record frontier: source-preserving movements and their target consequences can be calculated in the state coordinates. If $z_0 \in O$, every residual state direction has $d\alpha = 0$. The source-fiber criterion also applies directly to any declared measurement map; the quotient supplies reduced coordinates when its conditions hold.

COROLLARY 2 (State-space record bound). *Under the regularity of* (18),

$$d_{O\to T}(x) \le \dim\ker D\widetilde{Y}_O(s) = \dim S - \operatorname{rank} D\widetilde{Y}_O(s) \le r_M. \tag{19}$$

When the reference environment belongs to O, residual movements lie in the interaction-coordinate tangent space, so $d_{O\to T}(x) \le r$. Thus the common-state representation bounds the number of smooth scalar fields that can remain relevant for every declared target family, while (18) *gives the target-specific count.*

The local dimension is a recovery requirement. Turning it into a feasible recording policy additionally requires the measurement menu and a loss for the intended use. The bilinear benchmark below makes these inputs explicit and supplies the marginal value used by the subsequent menu search.

4.2. One residual map, exact and approximate frontiers

In the exact bilinear system, let $q \in \mathbb{R}^n$ be the exposure vector and let $V_O \in \mathbb{R}^{n\times K}$ and $V_T \in \mathbb{R}^{n\times J}$ collect the source and target loadings as columns. Write $\mathcal{V}_O = \operatorname{im} V_O$, $\mathcal{V}_T = \operatorname{im} V_T$, and

$$R_{O,T} = P_{\mathcal{V}_O^\perp} V_T, \qquad D(O,T) = \dim(\mathcal{V}_O + \mathcal{V}_T) - \dim \mathcal{V}_O. \tag{20}$$

For $q = q_y + w$ on a source fiber, $w \in \ker V_O'$ and the target innovation is $V_T' w = R_{O,T}' w$. This residual map is the bilinear form of the variation in (3).

To attach average loss, let $\mu_q = \mathbb{E}q$, $\Sigma_q = \operatorname{Var}(q)$ with finite second moments, and $S_q = \operatorname{im}\Sigma_q$. The covariance after best affine prediction from $V_O' q$ is

$$\Sigma_{q|O} = \Sigma_q - \Sigma_q V_O (V_O' \Sigma_q V_O)^\dagger V_O' \Sigma_q. \tag{21}$$

Fix a target metric $W_T \succ 0$ and define

$$K_{O,T}^{(q)} = \Sigma_{q|O}^{1/2} R_{O,T} W_T R_{O,T}' \Sigma_{q|O}^{1/2}, \qquad M_T = V_T W_T V_T'. \tag{22}$$

The exposure distribution and target metric supply the averaging and units of loss. The subscript in $\Sigma_{q|O}$ denotes the residual from best affine prediction. Under Gaussian exposures it also equals the conditional covariance given the source record; finite second moments suffice for the affine-loss calculation used here. Let $\lambda_1 \ge \cdots \ge \lambda_n \ge 0$ denote the eigenvalues of this operator, and write $\widetilde{\mathcal{V}}_A = P_{S_q}\mathcal{V}_A$.

THEOREM 3 (Sharp target-relative measurement frontier). *For the exact bilinear measurement system, the residual map $R_{O,T}$ gives the following frontiers.*

(i) Exact recovery. On $\mathbb{R}^n$, exact target recovery holds precisely when $R_{O,T} = 0$. The minimum dimension of freely chosen linear certificates is

$$k_{\text{exact}} = D(O,T) = \operatorname{rank} R_{O,T}. \tag{23}$$

A basis of $\operatorname{im} R_{O,T}$ gives completing exposure projections.

(ii) Worst-case loss. For scalar target $t'q$ and exposure ball $\mathcal{U} = B\mathbb{B}_2^n$, let $\operatorname{rad}(A) = \inf_c \sup_{a\in A} |a - c|$. With the supremum over attained source records,

$$\inf_f \sup_{q\in\mathcal{U}} |t'q - f(V_O'q)| = \sup_y \operatorname{rad} \mathcal{R}_{O\to t}(y;\mathcal{U}) = B\|P_{\mathcal{V}_O^\perp} t\|_2. \tag{24}$$

(iii) Average loss and information repair. With k freely chosen linear certificates and an affine decoder, minimum target-weighted mean-square loss equals

$$L_k^*(O,T) = \sum_{j>k} \lambda_j, \qquad \operatorname{rank} K_{O,T}^{(q)} = \dim(\widetilde{\mathcal{V}}_O + \widetilde{\mathcal{V}}_T) - \dim \widetilde{\mathcal{V}}_O. \tag{25}$$

This rank is the certificate count for zero affine-prediction loss and for exact affine recovery on the exposure envelope $\mu_q + S_q$.

(iv) Value of an added record. Put $\Sigma_O = \Sigma_{q|O}$. An added measurement $a'q$ with $a'\Sigma_O a > 0$ changes the residual covariance and target loss by

$$\Sigma_{O,a} = \Sigma_O - \frac{\Sigma_O a a' \Sigma_O}{a'\Sigma_O a}, \tag{26}$$

$$VM(a \mid O,T) = \frac{a'\Sigma_O M_T \Sigma_O a}{a'\Sigma_O a}. \tag{27}$$

Its value is zero when $a'\Sigma_O a = 0$. A certificate chooses a from the permitted object-side measurements; an environment record restricts a to a feasible loading $v(z)$.

PROOF OF THEOREM 3. For clause (i), differences of objects on a source fiber are exactly $\ker V_O' = \mathcal{V}_O^\perp$. Target variation on the fiber is the image of $V_T'|_{\mathcal{V}_O^\perp}$, whose rank is $\operatorname{rank} R_{O,T} = \dim(\mathcal{V}_O + \mathcal{V}_T) - \dim \mathcal{V}_O$. Exact transport requires this image to be zero. If k additional linear measurements recover the target, their restriction to the source kernel must separate this target image, so $k \geq \operatorname{rank} R_{O,T}$. Projections on an orthonormal basis of $\operatorname{im} R_{O,T}$ attain the bound.

For clause (ii), define q_y to be the minimum-norm solution of $V_O'q = y$. Every point in the ball on this fiber is $q_y + w$ with $w \in \mathcal{V}_O^\perp$ and $\|w\|^2 \leq B^2 - \|q_y\|^2$. The scalar residual-set radius is therefore

$$\operatorname{rad}\mathcal{R}_{O\to t}(y;B\mathbb{B}_2^n)=\sqrt{B^2-\|q_y\|^2}\,\|P_{\mathcal{V}_O^\perp}t\|_2. \tag{28}$$

The midpoint of the target interval minimizes the largest error on each fiber. Taking the supremum over attained records, including $y=0$, gives (24).

For clause (iii), factor $\Sigma_q=LL'$ with $L\in\mathbb{R}^{n\times s}$ of full column rank, $s=\dim S_q$. Let P_O project onto the row span of $V_O'L$. Orthogonal projection of centered exposures gives

$$\Sigma_{q|O}=L(I-P_O)L'. \tag{29}$$

Consequently $\Sigma_{q|O}V_O=0$ and $\Sigma_{q|O}^{1/2}R_{O,T}=\Sigma_{q|O}^{1/2}V_T$. The rank of the latter equals the rank increment from appending $V_T'L$ to $V_O'L$. Restriction of L' to S_q is an isomorphism, so this increment is the support-relative dimension in (25). If $\widetilde{\mathcal{V}}_T=S_q$, the increment is $s-\dim\widetilde{\mathcal{V}}_O=\operatorname{rank}\Sigma_{q|O}$, proving (31).

Write the centered source-residual exposure as $B\zeta$, where $BB'=\Sigma_{q|O}$ and $\mathbb{E}[\zeta\zeta']=I$ on its range. Freely chosen linear certificates select a subspace of ζ of dimension at most k. If P projects onto that subspace, the remaining weighted linear-prediction loss is

$$\operatorname{tr}[(I-P)B'M_TB(I-P)]=\operatorname{tr}[(I-P)B'M_TB]. \tag{30}$$

The Ky Fan principle maximizes the removed trace by selecting the leading k eigenvectors. The positive eigenvalues of $B'M_TB$ equal those of $K_{O,T}^{(q)}$. This gives the eigenvalue tail and its zero-loss certificate count. On μ_q+S_q, zero residual rank also means exact linear recovery everywhere, by the source-fiber argument.

For clause (iv), regress the centered residual exposure on $a'q$. Its covariance reduction is $\Sigma_O aa'\Sigma_O/(a'\Sigma_O a)$ when the denominator is positive. Taking the trace against M_T gives (27). If $a'\Sigma_O a=0$, positive semidefiniteness implies $\Sigma_O a=0$, and the added measurement has zero residual variation. □

Each clause uses the same residual target map with a different set of declared primitives. Exact rank uses the full exposure region, worst-case loss uses a bounded set, and average loss uses exposure moments together with a target metric. Certificates and environment records share the marginal-value formula in (27); their distinct action sets specify which measurement can be added. In this way, the residual geometry becomes a record-design problem.

Projection and rank identities give the exact count. The average-loss tail uses the Ky Fan variational principle and the geometry of reduced-rank prediction (Fan 1949; Tso 1981). Rank-one covariance updates and target-weighted design connect the marginal value to the optimal-design tradition (Elfving 1952; Kiefer and Wolfowitz 1960; Attia et al.

2018). For Gaussian exposures, the average-loss criterion equals the zero-noise limit of goal-oriented A-optimal design, including singular exposure supports. The exact bilinear assumptions govern these formulas; the nonlinear quotient supplies the local residual dimension through (18). Extending the finite-loss formulas to a nonlinear neighborhood also requires control of the approximation remainder over that neighborhood.

When target loadings span the centered exposure support, (25) becomes

$$D_{S_q}(O,T) = \operatorname{rank}\Sigma_{q|O} = \dim S_q - \dim \widetilde{\mathcal{V}}_O. \tag{31}$$

This is the accounting used in F-MAP: a six-dimensional centered basket space, four independent source valuations, and two remaining target-relevant directions. The dimension statement concerns linear records on the affine exposure envelope. Held-out evaluation checks the trained record and decoder on observed baskets.

At a tolerated average loss $\epsilon \geq 0$, the effective certificate count is

$$D_{\epsilon,W_T}(O,T) = \min\{k : \textstyle\sum_{j>k}\lambda_j \leq \epsilon\}. \tag{32}$$

The target metric belongs to the economic loss specification. A zero eigenvalue identifies a direction with zero contribution to this affine-loss criterion on the declared exposure envelope. Small positive eigenvalues value approximate recovery; their size depends on the exposure moments and target weights. Under $V_T \mapsto V_T A$, the congruent transformation $W_T \mapsto A^{-1} W_T A^{-\top}$ leaves the residual-loss operator unchanged. Changing the coordinate units while keeping its numerical metric fixed changes the loss being measured.

4.3. A feasible menu and its selection criterion

A certificate must be available at the recording date. Let $P(x)$ collect the permitted pre-target object inputs.

PROPOSITION 2 (Admissible information). *The combined input (Y_O, P) permits target recovery precisely when*

$$(Y_O(x), P(x)) = (Y_O(x'), P(x')) \implies Y_T(x) = Y_T(x'). \tag{33}$$

Under constant ranks of the combined source and source–target maps, its local smooth counterpart is

$$\ker DY_O \cap \ker DP \subseteq \ker DY_T. \tag{34}$$

Every certificate $C = c \circ P$ enabling smooth local recovery satisfies this necessary kernel inclusion.

When $P(q) = q$, the institution can form a quantity projection when it constructs the record and retain that scalar for future targets. A system with a restricted input P faces the gate in Proposition 2. The unrestricted certificate count in Theorem 3 is the benchmark for a record system that can form the corresponding exposure projections.

Let $\mathcal{Z}_0$ be a finite menu and K an externally supplied capacity. The menu-constrained policy optimum under average loss is

$$L^*_{\text{menu}}(K) = \min_{\substack{O \subseteq \mathcal{Z}_0 \\ |O| \leq K}} \operatorname{tr}(M_T \Sigma_{q|O}). \tag{35}$$

The integer K is the institution's record capacity. The statistical dimension $\widehat{r}$ summarizes an estimated spectrum and is a separate calibration diagnostic. The empirical frontier evaluates every supplied value of K and reports $\widehat{r}$ alongside it.

At each step, the implemented greedy target-directed policy adds the feasible environment maximizing (27). This is a conditional one-step improvement; the joint subset objective in (35) is the benchmark for comparing complete source sets. The policy optimum in (35) minimizes the declared calibration objective over all feasible subsets; held-out policy effects evaluate the subsets chosen by the implemented rule. Finite-menu enumeration and one-exchange search measure the search component. With environment costs $c_E(z)$ and object-field costs $c_C(a)$, the institution solves

$$\min_{O,A} \operatorname{tr}(M_T \Sigma_{q|O,A}) + \lambda \left\{ \sum_{z \in O} c_E(z) + \sum_{a \in A} c_C(a) \right\}, \tag{36}$$

over its feasible environment and certificate menus, where $\Sigma_{q|O,A}$ is the exposure covariance after projection on both record types. Conditional on the current record, a candidate field enters the one-step solution when its marginal value in (27) exceeds its loss-unit cost λc. The empirical capacity–loss schedule supplies the loss side of this comparison for a common per-field cost; application-specific collection, storage, and compliance costs supply the other side. The threshold is a one-step comparison conditional on the current record. Complete-set selection continues to use the objective over the feasible menu.

For orientation, classical subspace design allows any K-dimensional source span W. With target-direction second moment $\Omega_T = \mathbb{E}[tt']$, the minimum *average squared minimax error* over targets on the exposure ball is

$$\min_{\dim W \le K} B^2 \mathbb{E}\|P_{W^\perp} t\|_2^2 = B^2 \sum_{j>K} \lambda_j(\Omega_T). \tag{37}$$

This is the leading-eigenspace benchmark. It averages the squared worst-case error for each target; (25) averages prediction error over random exposures. The feasible environment menu in (35) and the declared loss specify the institution's actual choice.

5. Evidence from food baskets and cash flows

5.1. The F-MAP retention laboratory

The exercise uses the ten metropolitan-area series in USDA Food-at-Home Monthly Area Prices, which reports purchase dollars, purchase grams, and unit values for 90 expenditure food product groups during 2012–2018 (U.S. Department of Agriculture, Economic Research Service 2024). ERS constructs the published series from proprietary Circana scanner inputs and store-level weights. The analysis aggregates dollars and grams into seven consumption groups and forms each group unit value from these totals. Objects are normalized area–month quantity vectors q_i, environments are group price vectors p_e, and the oracle valuation is

$$m(q_i, p_e) = p_e' q_i. \tag{38}$$

Quantities use 100-gram units and sum to 1,000 per basket. The adding-up restriction places centered baskets in a six-dimensional hyperplane. The full six-dimensional rank of the calibration covariance is checked in the exercise below. The categories are dairy, fruit, grains, meat and protein foods, other foods, prepared meals, and vegetables.

The primary retention task uses the first tier of USDA's published food-group hierarchy (U.S. Department of Agriculture, Economic Research Service 2024). Its seven fields describe basket composition in consumption categories: an added scalar can report a weighted shift among fruit, grains, vegetables, and the other named groups. The composition-index dictionary below implements that use. Group unit values incorporate the within-group purchase mix in each environment, so the target is expenditure at that reporting resolution. Product mix and quality composition enter the represented environment together with transaction prices. Revaluation holds the normalized basket fixed. Its expenditure loss is an accounting target; a cost-of-living interpretation would additionally specify household substitution and preferences. The normalized system has six centered support directions, allowing the complete capacity and certificate frontier to be evaluated from zero through exact completion. The 34- and 90-group systems retain progressively finer product composition and define their own revaluation maps and capacity comparisons. Their reference capacities cover 8/33 and 13/89 of their centered support dimensions, compared with 4/6 in the primary system; the reported effects therefore pair each resolution with its declared capacity.

Training covers 2012–2015 with 480 area–month environments, validation uses the 120 environments in 2016, and evaluation uses the 240 environments in 2017–2018. Every evaluation basket is valued at every evaluation price vector, producing 57,600 pairs from the same 240 area–month identifiers. Each identifier occurs in both roles. Training quantities determine exposure moments; validation prices determine source selection and certificate order. Held-out quantities and prices determine evaluation losses.

Capacity is supplied as $K \in \{1, \ldots, 6\}$. The four-record point is the reference comparison, and every budget is displayed. Separately, a training-price rule with a 95-percent explained-variation floor and a 0.10 relative eigengap yields $\widehat{r} = 4$. This rule summarizes calibration-price variation. The six-dimensional basket support, the four-field budget, and the two completing directions answer separate questions: which compositions occur, how many fields are retained, and which target variations those fields leave unresolved. The explained-variation rule supplies a diagnostic for the chosen calibration sample; the exact completion count follows from the source and target loadings on the basket support.

Let P_O contain one retained source price vector per row, so $P_O = V_O'$ under the residual-information notation. The affine decoder is

$$\widehat{m}(q_i, p_t; O) = p_t'\mu_q + p_t'\Sigma_q P_O'(P_O\Sigma_q P_O')^{\dagger} P_O(q_i - \mu_q). \tag{39}$$

Its mean-square linear-projection loss is $p_t'\Sigma_{q|O}p_t$. A support-root singular-value decomposition evaluates this decoder numerically. Quantities enter calibration and the oracle, while the decoder receives the K source values and any retained certificates. The design question is which records carry information into later revaluation. The F-MAP series supplies a laboratory for that counterfactual reporting decision; the estimand concerns the retained field menu and downstream decoder.

5.2. Evaluating a selected retention rule

Fix calibration data $\mathcal{S}_0$, a capacity K, a decoder class, and a target specification. Policy a chooses a feasible source set O_a of cardinality K. For held-out pair (i, t), let $\widehat{m}_{it}(a)$ use the retained source values and the fitted decoder, and define

$$\ell_{it}(a) = [m(x_i, z_t) - \widehat{m}_{it}(a)]^2, \qquad \tau_{a:b} = \frac{1}{N_x N_t} \sum_{i,t} [\ell_{it}(b) - \ell_{it}(a)]. \tag{40}$$

The oracle value $m(x_i, z_t)$ and the trained policy maps give both losses for every evaluation pair. Thus $\tau_{a:b}$ is a directly observed finite-array loss contrast between retained-information rules. The object, target environment, calibration data, capacity, and decoder class are held fixed. This intervention concerns the information supplied to a future revaluation task.

For random source selection A, the comparator is $\mathbb{E}_A \ell_{it}(A)$. Uniform draws estimate this policy expectation. The random-policy tail locates the chosen rule in the declared distribution over feasible source sets, with the finite-draw correction. It is a policy-menu reference probability. Repeated-population references condition on the trained rules. Two-way clustering treats object and target dimensions separately, a node jackknife removes each area–month from both roles, and the forward cell assigns distinct calendar years to the two roles. The finite evaluation-array effect is computed directly.

The main results report fixed-policy intervals, loading sensitivity, and selection-aware sensitivity separately. The fixed-policy intervals condition on the rules selected before evaluation and use the stated repeated-population reference. The loading exercise perturbs published unit values under a declared covariance specification and reselects the records. Its percentile ranges summarize those perturbations. Online Appendix E.5 gives the common-multiplier construction, and Online Appendix E.7 details the loading inputs and covariance alternatives.

For confidence-protected capacity, the simultaneous upper bounds apply to capacities one through five. The six-field endpoint uses exact completion in the stated seven-group bilinear system. Numerical recovery is assessed at the reported tolerance. Capacity selection consequently combines the simultaneous bound with a separately justified algebraic endpoint.

What held-out loss measures. The affine decoder admits an exact accounting on a complete basket–price array. Hold its fitted μ_q, Σ_q and source set O fixed, and define

$$B_O = I - \Sigma_q V_O (V_O' \Sigma_q V_O)^{\dagger} V_O', \qquad M_{\text{eval}} = \frac{1}{N_t} \sum_t p_t p_t'. \tag{41}$$

Let $\bar{q}_E = N_x^{-1} \sum_i q_i$ and $\Sigma_E = N_x^{-1} \sum_i (q_i - \bar{q}_E)(q_i - \bar{q}_E)'$ be the evaluation moments, and set $\delta_E = \bar{q}_E - \mu_q$. Averaging the squared error $[\,p_t' B_O (q_i - \mu_q)]^2$ gives

$$\bar{\ell}_E(O) = \operatorname{tr}(M_{\text{eval}} B_O \Sigma_E B_O') + \delta_E' B_O' M_{\text{eval}} B_O \delta_E. \tag{42}$$

Appendix E.4 gives the expansion. The first term values residual composition variation; the second values a shift in mean composition. Later target prices weight both. The identity concerns the finite Cartesian evaluation array. Sampling assumptions enter the interpretation of intervals around that loss. It explains why the population attached to a calibration optimum matters for subsequent loss. Appending certificate loadings to V_O gives the same calculation for the augmented record.

5.3. Two residual dimensions and two completing certificates

At $K = 4$, greedy source valuations have rank four on the six-dimensional centered basket support. Evaluation target prices span all six support directions. Equation (31) therefore predicts two target-relevant missing coordinates. The residual covariance has rank two. The first two certificates are formed from the validation target-loss operator on this calibration residual space.

TABLE 1. Residual information and held-out revaluation

Retained information	Certificates	RMSE	MSE	MSE removed
Four source values	0	0.153629	0.023602	—
First residual certificate	1	0.075711	0.005732	75.7%
Two residual certificates	2	5.1×10^{-14}	2.6×10^{-27}	100.0%

The centered basket support has dimension six, the source-loading rank is four, and the target-loading rank is six. Certificate directions use training quantities and validation prices. The table evaluates all 57,600 held-out pairs. Numerical recovery is assessed against an absolute value-error tolerance of 10^{-9}.

One certificate removes 75.7 percent of remaining MSE; two recover every held-out target value to numerical precision. These certificates record quantity-composition directions that the four source expenditure numbers leave unresolved. The affine-support identity predicts their count; the evaluation checks the trained directions and decoder on later baskets and prices.

Each certificate can be stored as one composition index. Let $s_{ig} = q_{ig}/1,000$ be group g's share of basket mass. The certificate directions are recoded as

$$C_j(q_i) = 100 \sum_{g=1}^{7} w_{jg} s_{ig}, \qquad \sum_{g=1}^{7} w_{jg} = 0, \qquad \max_g |w_{jg}| = 1. \tag{43}$$

These scalar fields use a common weight dictionary. Moving one percentage point of basket mass from vegetables to fruit changes the first index by 1.694 points; moving that share from vegetables to grains changes the second by 1.519 points. This affine

TABLE 2. Held-out loss at each record capacity

Budget K	Greedy RMSE	Dispersed RMSE	Random RMSE	Certificates $6-K$
1	0.679937	0.679937	1.045165	5
2	0.419336	0.649119	0.645690	4
3	0.260297	0.583402	0.439927	3
4	0.153629	0.288628	0.293871	2
5	0.075742	0.153754	0.175604	1
6	2.0×10^{-12}	1.5×10^{-13}	4.2×10^{-12}	0

Each capacity is supplied independently of the training-price dimension diagnostic. The random column is the square root of expected MSE across 4,999 uniform source sets at that capacity. The last column records the completing certificate count for the greedy and dispersed source rules. Every row uses the same calendar split and evaluation array.

recoding preserves the information in the certificates.[7] Revaluation from four expenditure fields and two composition-index fields closes the two-dimensional deficit to numerical precision.

The count is reproduced across the capacity frontier. For both greedy and dispersed source sets, the completing certificate count equals $6 - K$ at every $K = 1, \ldots, 6$. The counts agree at relative rank tolerances 10^{-8}, 10^{-10}, and 10^{-12}. Ten leave-one-metropolitan-area-out calibrations repeat the four-record exercise: training and validation exclude one area, and evaluation uses that area's 24 later baskets against all 240 later targets. Each calibration again requires two completing certificates. Closure errors and calibration-sample comparisons confirm the count.

5.4. The capacity frontier and environment choice

At each capacity, greedy target-directed acquisition adds the training environment giving the largest reduction in validation target loss. Dispersed acquisition shares the first environment and then selects the farthest point in standardized training-price coordinates. Uniform acquisition draws K distinct training environments. The greedy target-directed rule implements the marginal value in (27) on the feasible price menu. The menu-constrained policy optimum in (35) is the calibration benchmark; the tables evaluate the implemented greedy policy on held-out values.

At $K = 5$, greedy and dispersed RMSE are 0.0757 and 0.1538. At $K = 6$, source values span the centered basket support and the deterministic rules recover the target array to numerical precision. The record frontier and the certificate frontier reach the same six-coordinate envelope through environment loadings or retained quantity projections.

[7] Any invertible change of basis within the two-certificate span preserves exact closure. The displayed weights choose an interpretable basis within that span.

TABLE 3. Simultaneous inference and confidence-protected capacity

K	Exchange MSE	RMSE (bp)	Effect	Simultaneous 95% interval	UCB RMSE (bp)
1	0.462314	20.60	0.00000	[0.00000, 0.00000]	22.77
2	0.157026	12.00	0.26433	[0.17685, 0.35181]	13.38
3	0.067946	7.90	0.27241	[0.17894, 0.36589]	9.02
4	0.018594	4.13	0.06471	[0.03668, 0.09274]	4.74
5	0.004691	2.07	0.01895	[0.01165, 0.02624]	2.48
6	0.000000	0.00	0.00000	[0.00000, 0.00000]	0.00

Effects are dispersed loss minus exchange-policy loss. A common 4,999-draw Gaussian multiplier acts on delete-one-area–month pseudovalues for the effect and exchange MSE at capacities one through five. The displayed intervals and MSE upper bounds use the resulting maximum studentized critical value of 2.676. Capacity six is exact algebraic closure. RMSE is measured in basis points of the $330.145 calibration benchmark expenditure.

At four records, greedy acquisition lowers the complete-array MSE from 0.083306 to 0.023602, a 71.7-percent reduction. Of 4,999 uniform source-set draws, 92 attain loss weakly below greedy loss, giving the corrected tail $(92 + 1)/5,000 = 0.0186$. The 57,600 evaluation cells reuse the same 240 area–month observations in basket and price roles. A delete-one-node jackknife therefore removes each area–month from both roles; its effect interval is $[0.03901, 0.08040]$.[8] The simultaneous comparison applies this common-node construction across capacities. A row–column calculation and a disjoint-calendar evaluation provide separate references. The finite-array loss contrast itself is observed directly.

The four-record exchange policy reduces RMSE from 0.2886 under dispersed acquisition to 0.1364. The difference is $0.1523 per normalized 100-kilogram basket, equal to 4.612 basis points of the $330.145 calibration benchmark expenditure. MSE falls by 0.064712 squared dollars per basket–target pair. For N future revaluations, a cost-weighted loss criterion admits incremental search and implementation cost up to $N\kappa(0.064712)$, where κ converts squared-dollar loss into the institution's cost unit. Both rules retain four scalar fields, so the common per-field storage component cancels in this comparison.

The four-record effect has simultaneous interval $[0.03668, 0.09274]$ around the estimate 0.06471. Exchange RMSE is 4.13 basis points, with a simultaneous upper bound of 4.74. Under this fixed-rule, node-based reference, the smallest protected capacities are three records for a 10-basis-point ceiling, four for 5 basis points, five for 2.5 basis points, and six for 1 basis point. These statements condition on the validation-selected policies before the 2017–2018 evaluation.

The loading sensitivity analysis uses the public standard error attached to each weighted EFPG unit value. Crossing 499 draw-specific policy reselections with ten held-out

[8]The repeated-node reference conditions on trained policies. Serial and cross-area dependence require an additional block or sampling specification.

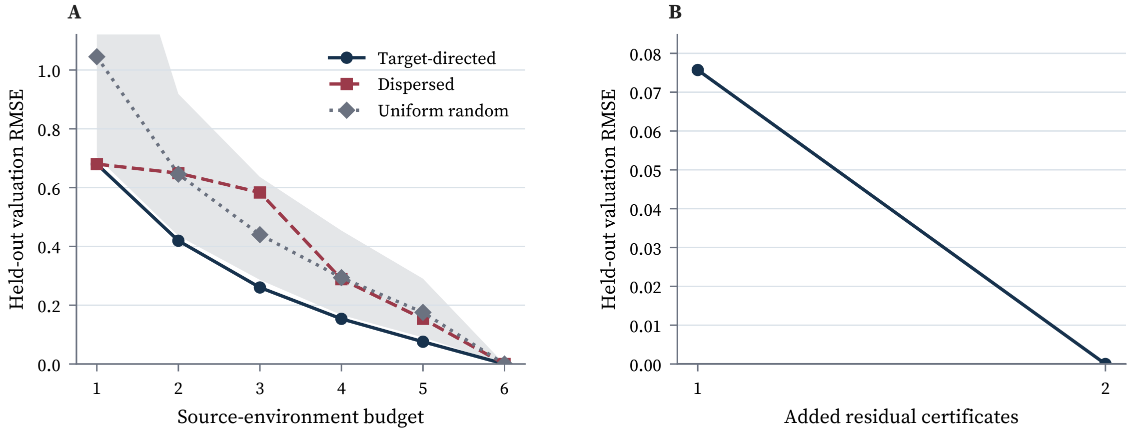

FIGURE 1. Capacity and residual certificates in held-out revaluation

Panel A varies retained source capacity from one to six environments. Curves report held-out valuation RMSE for target-directed, dispersed, and uniform-random source acquisition; the shaded region spans the fifth through ninety-fifth percentiles across uniform source-set draws. Panel B adds calibration-based residual certificates to four target-directed source values. Both panels evaluate the same 57,600 held-out revaluations. Capacity six and two added certificates close the corresponding finite arrays algebraically.

node perturbations gives 4,990 draws, a 75.9-percent median MSE reduction, and a 44.2–90.1 percentile range; all 4,990 loss effects are positive. Across five equicorrelation specifications for the unavailable EFPG covariance, the exchange median ranges from 72.8 to 76.1 percent and every 2.5th percentile exceeds 31 percent. The selection-aware construction carries the stated covariance scope.

Search and held-out performance. Greedy selection optimizes each additional record conditional on the current set. Exchange search revisits those choices. At four records, six accepted exchanges lower calibration MSE from 0.016848 to 0.014863 and held-out MSE from 0.023602 to 0.018594. The latter is a 77.7-percent reduction relative to dispersed acquisition. Exhaustive comparisons on specified smaller menus measure the remaining search gap. Full-menu, reduced-menu, and disjoint-calendar comparisons keep the calibration subset optimum, implemented search rule, and held-out loss as separate entries.

5.5. The marginal value of retained capacity

Let $L_{\text{eval}}(K)$ be held-out MSE along the greedy acquisition sequence. At $K = 0$, each target is valued using the training mean basket. One additional retained scalar saves

$$\Delta L_{\text{eval}}(K) = L_{\text{eval}}(K-1) - L_{\text{eval}}(K), \qquad K = 1, \dots, 6. \tag{44}$$

The calibration analogue equals the marginal record value in (27). Equation (44) evaluates that increment on subsequent baskets and targets.

TABLE 4. Marginal value of one retained scalar

K	MSE	Saved per scalar	Previous loss saved (%)	RMSE (basis points)
0	40.889540	—	—	193.687
1	0.462314	40.427227	98.9	20.595
2	0.175843	0.286471	62.0	12.702
3	0.067755	0.108088	61.5	7.884
4	0.023602	0.044153	65.2	4.653
5	0.005737	0.017865	75.7	2.294
6	4.0×10^{-24}	0.005737	100.0	6.1×10^{-11}

MSE and saved loss are in squared dollars for the normalized 100-kilogram basket. Each transition adds one source-value field. The percentage column divides saved loss by the preceding capacity's MSE. The last column is $10,000\sqrt{L_{\text{eval}}(K)}/\bar{v}_{\text{cal}}$, where $\bar{v}_{\text{cal}} = \bar{p}'_{\text{validation}}\mu_q$ = \$330.145 is the calibration benchmark expenditure. One basis point is 10^{-4} of that benchmark.

The fourth record saves 0.044153 in squared-dollar loss per basket–target pair, or 65.2 percent of the three-record loss. The fifth saves 0.017865, or 75.7 percent of the four-record loss. At four records, RMSE equals 4.653 basis points of benchmark expenditure; at five it equals 2.294. The sixth record completes the centered exposure support. This schedule expresses capacity in retained scalar fields and accuracy in valuation units. An institution can compare its entries with a supplied record cost and a chosen valuation-loss criterion.

5.6. Treasury exposure records

Treasury valuation applies the same calculation to dated promised cash flows c_i and discount curves $d(z_t)$:

$$m(c_i, z_t) = d(z_t)'c_i. \tag{45}$$

Federal Reserve curves determine the discount coordinates, and Treasury auction records determine the cash-flow schedules. The reference experiment retains one source valuation. Duration, duration plus convexity, and five key-rate measurements add cash-flow projections with exposure to target discount directions. These fields give economic names to the object-side certificates: each retained exposure summary supplies a scalar loading on the promised payment schedule.

The validation-optimal residual certificate minimizes the calibration loss over one additional linear field.[9] Its calibration MSE is 1.878, compared with 2.468 for duration. In the later evaluation, its RMSE is 2.234, while duration attains 0.789. Optimality refers

[9] Calibration uses 1990–2007 cash-flow moments and 2008–2015 target curves; the held-out comparison uses 2016–2025 cash flows and curves.

TABLE 5. Treasury certificate repair

Additional information	Held-out RMSE	MSE removed
Source valuation alone	14.563	—
Duration	0.789	99.71%
Duration and convexity	0.631	99.81%
Five key-rate certificates	0.212	99.98%
Validation-optimal residual certificate	2.234	97.65%

RMSE is in dollars per $100 face value. Reductions are relative to the one-source valuation decoder. Validation optimality uses 1990–2007 cash-flow moments and 2008–2015 target curves. Held-out losses use 2016–2025 cash flows and curves. Appendix F gives the construction and complete comparisons.

to the training-moment and validation-target criterion; the later cash-flow and discount populations determine the displayed ordering.

The Treasury records illustrate the approximate part of Theorem 3: each certificate removes target-weighted residual variation, with losses assessed in valuation units. Calendar-regime comparisons reject equality of the leading rank-one discount space with finite-draw $p = 0.0002$. This rotation evaluates the stability of a one-direction approximation. Certificate value is calculated against the actual target discount family, whose residual directions determine the repair task. The exact bilinear cash-flow map and its target-relative frontier remain the valuation primitives. Equality of leading one-dimensional spaces is the statistical null here; the full-support mixed kernel is the object in the representation theorem.

The active-space approximation test uses an eigengap gate and a dependence-preserving bootstrap. The Treasury construction applies the same tail-loss comparison to cash-flow records.

6. Conclusion

A record preserves the distinctions needed for its intended revaluations. The source-fiber criterion describes these distinctions directly. Under stable mixed kernels, a common interaction state and a reference valuation provide local coordinates in which to calculate them. Exact bilinearity then gives a residual target map whose rank counts completing fields and whose weighted spectrum values approximate recovery. The feasible menu turns these geometric benchmarks into a recording choice.

In the F-MAP exercise, four greedy source fields reduce held-out MSE by 71.7 percent relative to dispersed selection, and exchange refinement gives 77.7 percent. Two calibrated composition indices complete the remaining directions on the six-dimensional centered support. These comparisons concern a constructed accounting array and its fixed recording rules. The intervals and loading-reselection ranges retain their respective uncertainty references. The Treasury comparison places calibration optimality beside later valuation loss: duration performs better in the held-out sample than the field selected for the calibration criterion.

An institution using the frontier specifies its target family, permitted object inputs, available fields, and loss. A reference value carries levels, additional records carry the distinctions left open, and later environments determine how those distinctions are used. Collection costs, encoding precision, and behavioral responses can be included when the institution's question also concerns those margins.

References

Ingvild Almås, Orazio Attanasio, and Pamela Jervis. Presidential address: Economics and measurement: New measures to model decision making. *Econometrica*, 92(4):947–978, 2024. doi: 10.3982/ECTA21528.

Ahmed Attia, Alen Alexanderian, and Arvind K. Saibaba. Goal-oriented optimal design of experiments for large-scale bayesian linear inverse problems. *Inverse Problems*, 34(9):095009, 2018. doi: 10.1088/1361-6420/aad210.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24 (2):265–272, 1953. doi: 10.1214/aoms/1177729032.

Qiao Chen, Élise Arnaud, Ricardo Baptista, and Olivier Zahm. Coupled input–output dimension reduction: Application to goal-oriented bayesian experimental design and global sensitivity analysis. *SIAM Journal on Scientific Computing*, 47(5):A2403–A2430, 2025. doi: 10.1137/24M1670421.

Pierre-André Chiappori and Ivar Ekeland. Aggregation and market demand: An exterior differential calculus viewpoint. *Econometrica*, 67(6):1435–1457, 1999. doi: 10.1111/1468-0262.00085.

Paul G. Constantine, Eric Dow, and Qiqi Wang. Active subspace methods in theory and practice: Applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, 2014. doi: 10.1137/130916138.

W. Erwin Diewert. Exact and superlative index numbers. *Journal of Econometrics*, 4(2):115–145, 1976. doi: 10.1016/0304-4076(76)90009-9.

Gustav Elfving. Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, 23(2):255–262, 1952. doi: 10.1214/aoms/1177729442.

Ky Fan. On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proceedings of the National Academy of Sciences*, 35(11):652–655, 1949. doi: 10.1073/pnas.35.11.652.

Steven M. Goldman and Hirofumi Uzawa. A note on separability in demand analysis. *Econometrica*, 32(3):387–398, 1964. doi: 10.2307/1913043.

W. M. Gorman. Separable utility and aggregation. *Econometrica*, 27(3):469–481, 1959. doi: 10.2307/1909472.

W. M. Gorman. The structure of utility functions. *Review of Economic Studies*, 35(4):367–390, 1968. doi: 10.2307/2296766.

W. M. Gorman. Notes on Divisia indices. In Charles Blackorby and Anthony F. Shorrocks, editors, *Separability and Aggregation: The Collected Works of W. M. Gorman, Volume I*, pages 49–60. Oxford University Press, 1996. doi: 10.1093/0198285213.003.0004. Originally circulated as a 1970 typescript.

H. A. John Green. *Aggregation in Economic Analysis: An Introductory Survey*. Princeton University Press, Princeton, NJ, 1964.

Stefan Hoderlein and Arthur Lewbel. Regressor dimension reduction with economic constraints: The example of demand systems with many goods. *Econometric Theory*, 28(5):1087–1120, 2012. doi: 10.1017/S0266466612000412.

J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960. doi: 10.4153/CJM-1960-030-4.

Lap Chi Lau and Hong Zhou. A local search framework for experimental design. *SIAM Journal on Computing*, 51(4):900–951, 2022. doi: 10.1137/20M1386542.

Lucien Le Cam. Sufficiency and approximate sufficiency. *The Annals of Mathematical Statistics*, 35 (4):1419–1455, 1964. doi: 10.1214/AOMS/1177700372.

John M. Lee. *Introduction to Smooth Manifolds*. Springer, New York, 2 edition, 2013.

Kuang-Yao Lee, Bing Li, and Francesca Chiaromonte. A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Annals of Statistics*, 41(1):221–249, 2013. doi: 10.1214/12-AOS1071.

Wassily Leontief. A note on the interrelation of subsets of independent variables of a continuous function with continuous first derivatives. *Bulletin of the American Mathematical Society*, 53(4): 343–350, 1947a. doi: 10.1090/S0002-9904-1947-08796-6.

Wassily Leontief. Introduction to a theory of the internal structure of functional relationships. *Econometrica*, 15(4):361–373, 1947b. doi: 10.2307/1905335.

André Nataf. Sur la possibilité de construction de certains macromodèles. *Econometrica*, 16(3): 232–244, 1948. doi: 10.2307/1907277.

Susanne M. Schennach. Measurement systems. *Journal of Economic Literature*, 60(4):1223–1263, 2022. doi: 10.1257/jel.20211355.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pages 368–377, 1999. URL `https://arxiv.org/abs/physics/0004057`.

M. K.-S. Tso. Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(2):183–189, 1981. doi: 10.1111/j.2517-6161.1981.tb01169.x.

U.S. Department of Agriculture, Economic Research Service. Food-at-home monthly area prices, 2012–2018, 2024. URL `https://www.ers.usda.gov/data-products/food-at-home-monthly-area-prices/documentation`. Data product; documentation updated April 2026 and accessed August 2026.

Keyi Wu, Peng Chen, and Omar Ghattas. A fast and scalable computational framework for large-scale high-dimensional bayesian optimal experimental design. *SIAM/ASA Journal on Uncertainty Quantification*, 11(1):235–261, 2023. doi: 10.1137/21M1466499.

Olivier Zahm, Paul G. Constantine, Clémentine Prieur, and Youssef M. Marzouk. Gradient-based dimension reduction of multivariate vector-valued functions. *SIAM Journal on Scientific Computing*, 42(1):A534–A558, 2020. doi: 10.1137/18M1221837.

Shijie Zhong, Wanggang Shen, Tommie Catanach, and Xun Huan. Goal-oriented bayesian optimal experimental design for nonlinear models using markov chain monte carlo. *SIAM/ASA Journal on Uncertainty Quantification*, 14(1):19–47, 2026. doi: 10.1137/24M1649344.