Ordinal Complementarity*

Martín Almuzara[†] Martín García-Vázquez[‡]
Joseph Mullins[§]

This version: August 2026

Abstract

This paper introduces ordinal complementarity: a feature of stochastic production functions that is robust to monotonic transformations of inputs and output. It has two equivalent definitions: the first being supermodularity of the conditional survival function under a first-order stochastic dominance ordering, and the second being supermodularity of $\mathbb{E}[Q(Y)|X]$ for all non-decreasing valuations Q. Relative to existing conventions, this definition offers a strengthening with more robust policy implications. Ordinal complementarity is very strong. It is necessary and sufficient for a number of social planner problems to have monotone solutions for all Q, including those with arbitrarily strong redistributive preferences. It is testable — along with a number of related but weaker production properties — because it can be represented as a set of moment inequalities. An empirical application to the technology of cognitive skill formation for children finds different combinations of these properties at different ages, leading to new insights.

Keywords: complementarity, skill formation, supermodularity, stochastic dominance, optimal transport, moment inequalities.

*We thank Stéphane Bonhomme, Ismael Mourifié and Brian Rogers for insightful comments and discussions. We also thank seminar participants at UCLA, Western, and participants of the EWMES and MEA Structural Microeconomics for useful comments. Finally, we thank Maxwell Uphof for outstanding research assistance. The views expressed in this paper do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. All errors are our own.

[†]FRBNY: martin.almuzara@ny.frb.org

[‡]Washington University in St. Louis: garciavazquez@wustl.edu

[§]University of Minnesota: mullinsj@umn.edu

1 Introduction

This paper develops and empirically tests a notion of production complementarity (a higher value of one input raising the return to another) that is ordinal, meaning that it holds across all valuations of the output. An alternative definition, based on increasing differences in average production outcomes, has become the standard in the education and human capital literature. It has also become a key organizing concept and empirical focal point, because higher (or lower) returns to investment for individuals with higher initial conditions, such as current skills or genetic endowments, are suggestive of general policy implications (Cunha and Heckman, 2007). The robustness of these implications is illusory for the standard definition, however, because it is not consistent across equally plausible monotonic transformations of the outcome. The version of *ordinal complementarity* that this paper introduces strengthens the traditional definition to obtain more generalizable policy content: in two broad classes of planning problems it is both a necessary and sufficient condition for optimal allocations to be monotone. An empirical application to the production of cognitive skills in childhood, following Cunha et al. (2010), tests for ordinal complementarity and traces out its policy implications.

Consider a production process in which Y is the outcome of interest and X is a vector of inputs. In empirical settings production is inherently stochastic, so the production function is a conditional distribution $F_{Y|X}$, or equivalently a survival function $\overline{F}_{Y|X} = 1 - F_{Y|X}$. In search of a condition that guarantees, for any non-decreasing Q, that $E[Q(Y) \mid X]$ is supermodular in X, this paper defines ordinal complementarity as supermodularity of $\overline{F}_{Y|X}$ under a stochastic dominance ordering:

$$\frac{1}{2}\overline{F}_{Y|X_1 \wedge X_2} + \frac{1}{2}\overline{F}_{Y|X_1 \vee X_2} \succsim_{\text{FOSD}} \frac{1}{2}\overline{F}_{Y|X_1} + \frac{1}{2}\overline{F}_{Y|X_2}, \qquad \forall\, X_1, X_2, \tag{1}$$

where $\wedge$ and $\vee$ denote the componentwise minimum and maximum. The first set of results establishes the equivalence between this property and the supermodularity of $E[Q(Y) \mid X]$ for every non-decreasing Q (Athey, 1998). Supermodularity generally implies monotone allocations (Galichon, 2016; Topkis, 1978), and ordinal complementarity guarantees supermodularity for all Q, validating that this is a natural condition for generalizable policy conclusions. The paper therefore has tight connections to the literature on monotone comparative statics, where the central exercise is also to establish properties of objective functions that imply monotone solutions (Topkis, 1978). While ordinal results are available in that literature (Milgrom and Shannon, 1994), [1] two gaps separate that setting from the one studied here. First, the comparative statics literature works directly with objective functions, whereas in the class of problems considered here the production function defines only part of the objective. Second, the present interest is in bringing the concept to quantitative work, which means working with stochastic production functions rather than deterministic objectives. Athey (2002) bridges the second gap by considering stochastic optimization problems and giving conditions on distributions and objectives that generate cardinal or ordinal monotone comparative statics. This paper does the same for a specific class of problems suited to its conceptual and quantitative application, and for that class it establishes that supermodularity—rather than quasi-supermodularity or single crossing—is necessary and sufficient to guarantee monotone solutions.

The paper finds its motivating application in the education and human capital literatures, where complementarities are a common empirical focal point. In that application, Y is a vector of future skills or other human capital outcomes, θ is the current stock of skills, and I is an input. In quantitative work[2] the preferred approach specifies production as

$$Y = f(\theta, I) + \eta, \tag{2}$$

where η is a structural error. Overwhelmingly, the literature[3] equates complementarity with f exhibiting increasing differences in θ and I. While this property is of genuine interest, its content is limited to settings in which $E[Y \mid \theta, I]$ is inherently and exclusively the policy objective. Complementarity of the structural function f does not generalize to objectives of the form $E[Q(Y) \mid \theta, I]$, and it can be reversed for sufficiently concave transformations Q, which prohibits drawing any truly general policy conclusion from the shape of f.[4] Furthermore, Section 2.3 establishes under which conditions additive specifications exhibit ordinal complementarity.

[1] Milgrom and Shannon (1994) relax the conditions on objective functions to a pair of ordinal ones—quasi-supermodularity and a single-crossing property.

[2] See, for example, Cunha et al. (2010), Agostinelli and Wiswall (2016), Agostinelli and Wiswall (2025), and Caucutt et al. (2026).

[3] Some prominent examples include Cunha and Heckman (2007), Agostinelli and Wiswall (2025), Meghir et al. (2023), and Chetty et al. (2014). Other recent examples include Bolt et al. (2026); Gallegos and García (2024); Heckman et al. (2026); Hollenbach et al. (2026); Johnson and Jackson (2019); Kinsler (2016); Moroni et al. (2025); Muslimova et al. (2026). This is by no means an exhaustive list.

[4] A concrete illustration is the canonical CES technology $\theta_{t+1} = [a\,\theta_t^{\rho} + (1-a)I_t^{\rho}]^{1/\rho}$ with $\rho \in (0,1)$: its cross-partial is positive, yet the cross-partial of $\log \theta_{t+1}$ in (θ_t, I_t) is negative. The same technology delivers complementarity under one anchor and substitutability under a monotone relabelling of the very same latent output.

By contrast, Section 3 shows that when the production function $F_{Y|X}$ exhibits ordinal complementarity, one can draw policy conclusions that are consistent for *all* non-decreasing valuations Q. In particular, ordinal complementarity implies that optimal planner allocations of investment are non-decreasing in initial skills or endowments, with the converse holding under ordinal substitutability. These prescriptions are invariant to the choice of anchor Q and therefore hold no matter how strong the planner's redistributive motive.

A motivating example. To clarify the point, suppose $\theta \in \{\theta_L, \theta_H\}$ is cognitive skill, $I \in \{0, I_H\}$ is investment, and Y is cognitive skill next period, with

$$Y = \theta + I + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma^2). \tag{3}$$

Two education outcomes—Q_1 and Q_2, indicating completion of high school and of college—depend on cognitive skill through thresholds, $Q_1 = \mathbf{1}\{Y > c_1\}$ and $Q_2 = \mathbf{1}\{Y > c_2\}$ with $c_1 < c_2$. Education then follows an ordered probit. The latent technology (3) exhibits no complementarity in the traditional sense: $E[Y \mid \theta, I]$ has constant differences. Yet the implications for the two binary outcomes diverge. When θ_H lies well above the high-school cutoff c_1 and θ_L lies just below it, the nonlinearity of the normal CDF delivers

$$E[Q_1 \mid I_H, \theta_L] - E[Q_1 \mid 0, \theta_L] \; > \; E[Q_1 \mid I_H, \theta_H] - E[Q_1 \mid 0, \theta_H],$$

i.e. *decreasing* differences (substitutability) in high-school completion. Conversely, when θ_L lies far below the college cutoff c_2 and θ_H lies just below it,

$$E[Q_2 \mid I_H, \theta_H] - E[Q_2 \mid 0, \theta_H] \; > \; E[Q_2 \mid I_H, \theta_L] - E[Q_2 \mid 0, \theta_L],$$

i.e. *increasing* differences (complementarity) in college completion. Figure 1 displays both panels.

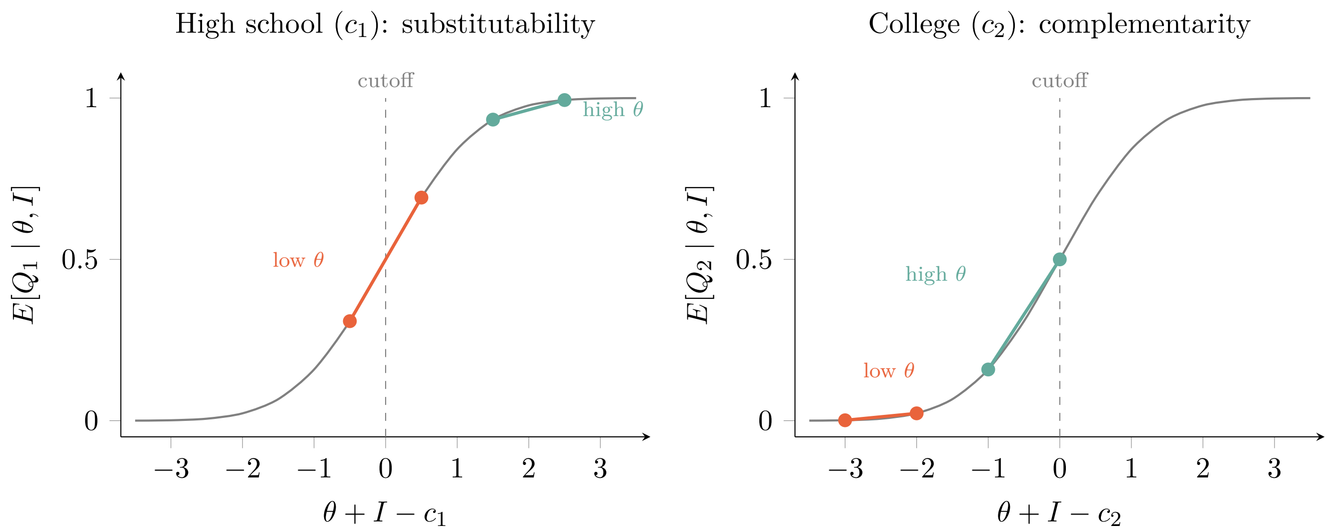

Figure 1: The same latent technology $Y = \theta + I + \eta$ yields substitutability for one binary outcome and complementarity for another. Each segment connects low to high investment at a fixed skill level; its slope is the return to investment. Left: high-school completion (c_1), where the high-skill type sits on the flat upper tail of Φ (ceiling effect) and the return to investment is larger for the low-skill type. Right: college completion (c_2), where the low-skill type sits on the flat lower tail (floor effect) and the return is larger for the high-skill type.

The story is intuitive. High-school completion is all but guaranteed for individuals who have already reached a high skill level, and the resulting ceiling on returns creates substitutability when high-school completion is the outcome of interest. College completion is nearly impossible for low-skill individuals regardless of investment, and the resulting floor creates complementarity. Substitutability or complementarity can therefore arise for outcomes that correspond to different nonlinear transformations of the same latent skill, despite there being no underlying complementarity in the technology of skill formation.

The example invites two comments. First, a researcher interested solely in maximizing Q_1 or solely in maximizing Q_2, and making complementarity statements exclusively with respect to that outcome, needs no additional generality; standard results then characterize how investment should be allocated across skill to maximize the average of that single outcome. This paper is at least partially predicated on the desire for properties that can generalize across multiple valuations or anchors. Second, the example shows that ordinal complementarity is a demanding property for production functions to satisfy in practice, as Section 2.3 shows for additive models.

Motivated by the second observation, one may ask whether there are weaker properties than ordinal complementarity that can still at least partially guarantee policy content. Section 2.2 proposes two weakenings based on how a technology may partition the space of Q-valuations into regions that are strictly super- or submodular. For example, while ordinal complementarity guarantees supermodular expectations for all Q, if this condition fails one may ask whether there exists *any* function Q such that $\mathbb{E}[Q(Y)|X]$ is sub- or supermodular. If such a Q cannot be found, then $F_{Y|X}$ is *ordinally non-substitutable* or *ordinally non-complementary*. The second weakening starts from the observation that concave Q functions can exacerbate substitution forces while convex Q functions do the opposite. If all convex Q lead to supermodular expectations, then $F_{Y|X}$ is *convex-ordinally complementary* and conversely *concave-ordinally substitutable* if every concave Q generates submodular expectations. These conditions both form strict supersets over the set of technologies that are ordinally complementary (or substitutable) and allow for more moderate generalizability.

For tractability, the paper's empirical application focuses on the case where output Y and inputs X have finite support. Section 4 advances the theory toward this application by outlining statistical tests for each of the above properties. In the main case of interest, Ordinal Complementarity implies a set of moment inequalities based on (1): the difference in differences of survival probabilities with respect to two comparison points in the domain of inputs X. The linearly independent set of these forms the basis of a test. In practice, the moment selection procedure of Romano et al. (2014) can help improve testing power when the number of inequalities under the null becomes large. Section 4 also develops an alternative test based on the observation that the set of ordinally complementary conditional distributions lives inside a convex cone, which means that one can employ the projection based test of Fang and Seo (2021).

In the economics of child development *dynamic complementarity* — defined as complementarity between inputs at different stages of childhood — has become an empirical focal point and an organizing concept since its introduction by Cunha and Heckman (2007). The concept is particularly compelling because it emphasizes the policy content of skill formation in childhood: if current inputs are complementary with past inputs then there is an equity-efficiency trade-off. If the converse holds then equity and efficiency align: returns to investment are higher for children with lower skills and more disadvantage. Because of this, among papers that seek to estimate the technology of skill formation, there is understandably an emphasis on whether such complementarities exist (Cunha et al., 2010; Cunha and Heckman, 2007; Heckman et al., 2026).[5]

[5]These exercises estimate and test the interaction between early and late investments implicitly by looking at the interaction between investment and current skills. Heckman et al. (2026) provide an exception: they

Sections 5 and 6 use panel data from the CNLSY[6] to ask whether cognitive skill formation exhibits any of the ordinal properties outlined in Section 2. At stake is the possibility of uncovering a strengthened version of dynamic complementarity with policy implications that generalize across different anchors of final cognitive skill. The CNLSY data have informed related parametric estimation exercises (Agostinelli and Wiswall, 2025; Cunha et al., 2010) and are a useful testing ground for these concepts. With the exception in this case of having a technology with discrete inputs and outputs, the modeling and estimation approach closely follows Cunha et al. (2010): a dynamic latent factor model where identification hinges on multiple measurements of both latent skills and latent investments (Bonhomme and Robin, 2009; Garcia-Vazquez, 2021; Hu and Schennach, 2008).

The results of the estimation and testing exercise yield a number of important and new insights. First, cognitive skill formation in early childhood does not appear to exhibit either ordinal complementarity or substitutability. This implies that knowledge of the complementarity properties of the skill formation technology are not sufficient to conclude whether optimal investment allocations will be increasing or decreasing in children's skills, and those conclusions will likely depend on the outcome of interest and societal preferences for redistribution. Furthermore, it is both ordinally non-substitutable and ordinally non-complementary: no increasing Q exists that can deliver sub- or supermodularity in expectations. Second, cognitive skill formation for school aged children appears to be very close to ordinally substitutable, meaning that the highest investments at school ages should be targeted to the lowest skill children no matter the anchor Q. Third, moment inequality tests cannot reject the null hypothesis that skill formation in early childhood exhibits both convex-ordinal complementarity and concave-ordinal substitutability.

Having analyzed interactions between investment and skills in the technology, Section 6 next analyzes the properties of *induced* transition equations when investment cannot be manipulated directly, but only indirectly through income, which shapes investment choices through estimated parental investment policies. Here the pattern of results is identical to the case with investment: ordinal non-substitubality and non-complementarity for young children, substitutability for older children. This suggests optimal reallocations of income should also be counter-monotone, targeting lower skilled children.[7]

directly test for the complementarity of current and lagged inputs using variation in the timing of entry to a randomized home visiting program in China.

[6]The Children of the 1979 cohort of the National Longitudinal Survey of Youth

[7]Of course this prescription is purely illustrative about the properties of the technology. These stylized optimization problems do not consider the perverse incentive effects of such an allocation.

A final set of counterfactuals solves two expository social planner problems to demonstrate the properties of the technology and look for consistent patterns in optimal allocations when the planner's objective employs different anchors Q. These two exercises seek to reallocate latent investment and household income in order to maximize a number of different adult outcomes (high school graduation, college graduation, employment, and earnings). Consistent with the testing evidence, which suggested that all anchors Q produce expectations that are neither super- nor submodular everywhere, optimal allocations of investment in early childhood are non-monotone for all choices of Q. Furthermore, and consistent with the finding that the technology is very close to ordinally substitutable at older ages, optimal policies are consistently countermonotone across all anchors.

Section 2 defines ordinal complementarity and establishes its invariance and equivalence properties, treats additive models and multiple outcomes, and develops some natural weakenings of ordinal complementarity. Section 3 draws out the efficiency content—static and dynamic—through optimal transport. Section 4 constructs a number of asymptotically valid tests for each of the conditions in Section 2. Section 5 describes identification and estimation of the technology of skill formation. Section 6 reports the empirical results and the counterfactuals. Section 7 concludes. Proofs not given in the text are collected in Appendix A.

2 Ordinal complementarity

This paper defines ordinal complementarity in pursuit of a condition on the conditional distribution $F_{Y|X}$ that guarantees $E[Q(Y) \mid X]$ is supermodular in X_1 and X_2 for any nondecreasing Q. Section 2.1 defines ordinal complementarity and establishes an equivalence class between it and ordinally invariant supermodularity of the expectation operator (in other words, it is both sufficient and necessary for this condition on expectations).

Section 2.2 then introduces two weaker notions, used in the application when fully ordinal conditions fail: ordinal non-substitutability, which only excludes substitutability, and the curvature-restricted convex complementarity and concave substitutability. Two further questions occupy the rest of the section. First, do commonly specified empirical production functions satisfy ordinal complementarity? Section 2.3 gives a mostly negative answer, ruling out large classes of specifications—a finding that shapes the paper's approach to estimation and testing in Section 5. Second, what extension is appropriate when Y is vector-valued? Section 2.4 offers two.

2.1 Setup and main definition

There are inputs $X = (X_1, \ldots, X_n)$ and an outcome Y, with $n \geq 2$. Inputs and outcome have an ordinal but not necessarily cardinal interpretation, and the relationship between X and Y may be stochastic. Let $\mathcal{X}_j$ and $\mathcal{Y}$ (subsets of $\mathbb{R}$) denote the supports of X_j and Y, and let the conditional CDF of Y given X be $F_{Y|X}(y \mid x) = \mathbb{P}(Y \leq y \mid X = x)$ for $y \in \mathcal{Y}$ and $x \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$. For vectors x, x', $x \wedge x'$ and $x \vee x'$ denote the elementwise minimum and maximum. Of course, if x and x' agree off the first two coordinates but are ordered in those coordinates as well, then $\{x \wedge x', x \vee x'\} = \{x, x'\}$, so supermodularity has content only outside the set of such pairs.

Our definition of ordinal complementarity involves the conditional survival function, which is given by$\overline{F}_{Y|X}(y \mid x) = 1 - F_{Y|X}(y \mid x)$.

Definition 1 (Ordinal complementarity)**.** X_1 and X_2 are ordinal complements in Y if and only if, for all $y \in \mathcal{Y}$ and all $x, x' \in \mathcal{X}$ such that $x_j = x'_j$ whenever $j > 2$,

$$F_{Y|X}(y \mid x \wedge x') + F_{Y|X}(y \mid x \vee x') \ \leq \ F_{Y|X}(y \mid x) + F_{Y|X}(y \mid x'). \tag{4}$$

In words, X_1 and X_2 are ordinal complements if the conditional survival function $\overline{F}_{Y|X}(y \mid x)$ is supermodular in (x_1, x_2), holding the remaining coordinates fixed—if X_1 and X_2 act as complements in raising the probability of high outcomes. Definition 1 extends to any pair of coordinates by reordering, and if every pair is ordinally complementary then $\overline{F}_{Y|X}(y \mid x)$ is supermodular in x.

Two properties justify Definition 1 in an ordinal setting. The first is invariance.

Proposition 1 (Invariance to monotone relabellings)**.** *For measurable, strictly increasing, real-valued functions G, g_1 and g_2, X_1 and X_2 are ordinal complements in Y if and only if $g_1(X_1)$ and $g_2(X_2)$ are ordinal complements in $G(Y)$.*

Proposition 1 (proved in Appendix A) is the precise sense in which the property is ordinal rather than cardinal: it addresses the anchoring problem of the introduction. Testing (4) is a test of complementarity in *any* system of measurement for the inputs and outcome, so a researcher need not take a stand on the cardinal scale of latent variables. The second property connects the definition to the full class of monotone valuations. Proposition 2 below establishes an equivalance class between $F_{Y|X}$ being ordinally supermodular and $\mathbb{E}[Q(Y)|X]$ being supermodular for all non-decreasing functions Q.

Proposition 2 (Supermodularity of expectations)**.** *X_1 and X_2 are ordinal complements in Y if and only if, for every measurable, increasing, integrable Q and all $x, x' \in \mathcal{X}$ with $x_j = x'_j$ whenever $j > 2$,*

$$E[Q(Y) \mid x \wedge x'] + E[Q(Y) \mid x \vee x'] \;\geq\; E[Q(Y) \mid x] + E[Q(Y) \mid x']. \tag{5}$$

Proposition 2 reflects what Athey (2002) calls the *closed convex cone property* of supermodular functions—the class of supermodular functions is closed under the non-negative combinations and limits that taking expectations performs—and is also implied by a result in Athey (1998); Appendix A gives a standalone proof for completeness. The proof rests on the link between supermodularity and first-order stochastic dominance: Definition 1 is exactly the statement that the average $\frac{1}{2}[F_{Y|X}(\cdot \mid x \wedge x') + F_{Y|X}(\cdot \mid x \vee x')]$ dominates $\frac{1}{2}[F_{Y|X}(\cdot \mid x) + F_{Y|X}(\cdot \mid x')]$, and integrating a non-decreasing Q against a dominated distribution preserves the inequality. An equivalent reading of ordinal complementarity is therefore complementarity in producing expected values of *every* monotone transformation of the outcome. Combined with Proposition 1, the conclusion (5) is moreover unaffected by any relabelling that respects the ordering. By contrast, approaches that rest on supermodularity of a production function are sensitive to the scale of X and Y, and so are unsuitable for the ordinal setting studied here.

2.2 Weaker notions of complementarity

The "for every Q" condition in Proposition 2 invites alternative classifications of the production technology based on classes of Q, either by asking only that substitutability be excluded for all Q or by looking for complementarity on restricted classes of Q. Both arise in the upcoming application.

The first weakening replaces the universal condition on Q with a negation: rather than asking that every valuation rate the inputs as complements, it asks only that no valuation rate them as substitutes.

Definition 2 (Ordinal non-substitutability)**.** $F_{Y|X}$ is *ordinally non-substitutable* in (X_1, X_2) if there is no measurable, increasing, integrable Q for which $E[Q(Y) \mid x]$ is submodular in (x_1, x_2); equivalently, for every such Q there is a pair x, x' with $x_j = x'_j$ whenever $j > 2$ at which

$$E[Q(Y) \mid x \wedge x'] + E[Q(Y) \mid x \vee x'] \;>\; E[Q(Y) \mid x] + E[Q(Y) \mid x'].$$

If such a Q exists, then the technology is *submodular-rationalizable.* Symmetrically, they are *ordinally non-complementary* if no such Q makes $E[Q(Y) \mid x]$ supermodular and *supermodular-rationalizable* if such a Q exists.

By Proposition 2, ordinal complementarity implies ordinal non-substitutability—if every valuation is supermodular then none is submodular—and the inclusion is strict: a technology whose local interactions change sign across the support can violate (5) for some Q yet admit no submodular valuation at all. The property does not certify complementarity; it only rules out substitutability, and it is the relevant hypothesis when the goal is to exclude substitute-like behaviour. Non-substitutability and non-complementarity are not nested: they can hold together, singly, or neither.

Definition 3 (Convex complementarity, concave substitutability)**.** X_1 and X_2 are *convex-ordinally complementary* in Y if $E[Q(Y) \mid x]$ is supermodular in (x_1, x_2) for every non-decreasing convex Q, and *concave-ordinally substitutable* if $E[Q(Y) \mid x]$ is submodular for every non-decreasing concave Q.

Since the convex (concave) valuations are a subclass of all non-decreasing ones, ordinal complementarity implies convex-ordinal complementarity and ordinal substitutability implies concave-ordinal substitutability, both strictly. Convex complementarity is the pertinent property when the outcomes that matter for policy load on the top of the distribution—as college completion does in the application, its value concentrated in the highest skill step.

2.3 Production functions with additive shocks

The motivating models of the human capital literature are additive, as in (2). When does ordinal complementarity hold in that class? In the absence of shocks, Definition 1 forces $Y = f(X)$ with f supermodular and $G(f(X))$ supermodular for every monotone G. No strictly increasing f can meet this requirement, which already signals that ordinal invariance is too strong in a deterministic world.[8]

With additive shocks $Y = f(X) + \varepsilon$, where $\varepsilon \perp X$ has CDF H and density h, the conditional CDF is $F_{Y|X}(y \mid x) = H(y - f(x))$ and ordinal complementarity reduces to a

[8]This is easy to see for the case of f strictly increasing and twice continuously differentiable, but true more generally for all strictly increasing f. If Y is measurable with respect to X then $Y = f(X)$ for some measurable f. Applying Proposition 2 with $Q(Y) = Y$ delivers supermodularity of f, and with $Q(Y) = G(Y)$ delivers supermodularity of $G(f(\cdot))$ for all monotone G. For twice continuously differentiable strictly increasing f and $G(y) = 1 - e^{-cy}$, supermodularity of $G(f)$ requires $c \le \partial^2_{x_1 x_2} f / (\partial_{x_1} f \, \partial_{x_2} f)$ everywhere, which fails for c large enough.

sign condition on a cross-partial derivative (Appendix A records two convenient criteria). Applied to standard shock distributions, with f strictly increasing and supermodular, they deliver the following.

(i) If $\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$, ordinal complementarity fails for every such f.

(ii) It also fails if $\varepsilon \sim \text{Uniform}(a, b)$, and more generally whenever the conditional support of Y given X depends on (x_1, x_2).

(iii) If $f(x_1, x_2) = A\, u(x_1)v(x_2)$ with $0 \leq u, v \leq 1$, and ε is Laplace, Logistic, or reflected Gumbel with scale σ, ordinal complementarity holds whenever $\sigma \geq A$.

The general pattern, made precise by Proposition 8 in the appendix, is that ordinal complementarity in additive models requires the shock to have a tail that is not too thin: it fails for any shock whose density has a score h'/h unbounded above (case (i)), and holds only once the shock is heavy enough to absorb the curvature of f (case (iii), illustrated in Example A). Two messages follow for empirical practice. First, the workhorse specifications—Cobb-Douglas or CES production with Gaussian shocks—do not satisfy ordinal complementarity for any supermodular f. Second, in case (iii) ordinal complementarity may hold even when there exists a monotone G with $G(f)$ not supermodular; all that is needed is one additive representation for which a criterion of Lemma 1 holds.

2.4 Multiple outcomes

Many applications feature a vector outcome $Y = (Y_1, \ldots, Y_r)$—for instance cognitive and non-cognitive skills. There is no unique extension of Definition 1; two are natural. Write $\overline{F}_{Y|X}(y \mid x) = \mathbb{P}(Y \geq y \mid X = x)$ for the joint survival function.

Definition 4 (Multivariate ordinal complementarity)**.** X_1 and X_2 are ordinal complements in Y in the *weak sense* if, for all y and all x, x' with $x_j = x'_j$ for $j > 2$,

$$\overline{F}_{Y|X}(y \mid x \wedge x') + \overline{F}_{Y|X}(y \mid x \vee x') \geq \overline{F}_{Y|X}(y \mid x) + \overline{F}_{Y|X}(y \mid x').$$

They are ordinal complements in the *strong sense* if the same inequality holds with $\overline{F}_{Y|X}(\cdot \mid \cdot)$ replaced by $\mathbb{P}(Y \in V \mid \cdot)$ for every upper set $V \subseteq \mathbb{R}^r$.

Since the orthant $\{y' \geq y\}$ is an upper set, the strong sense implies the weak sense; the two coincide when $r = 1$ and reduce to Definition 1. They correspond to different valuation classes in the supermodularity-of-expectations equivalence—the weak sense to products of monotone functions of each coordinate, the strong sense to every monotone vector valuation—and both remain testable whenever the conditional distributions are identified. Proposition 9 (Appendix A) records the equivalences and a sufficient condition for the strong sense: that the conditional copula of Y given X not depend on (x_1, x_2).

3 The policy content of ordinal complementarity

Ordinal complementarity has almost immediate policy content because of the role played by supermodularity in monotone comparative statics and optimal transport (Galichon, 2016; Topkis, 1978). This section formalizes this content by posing two social planning problems and showing that ordinal complementarity makes assortative allocations optimal for *every* non-decreasing valuation of the outcome, and is the weakest property that does so: it is necessary and sufficient for this guarantee in both problems. Throughout, X_1 and X_2 are the two inputs and Y the outcome, with planner valuation $q(x_1, x_2) = E[Q(Y) \mid X_1 = x_1, X_2 = x_2]$; in the application X_1 is current skill, X_2 investment, and Y future skill. Section 3.1 establishes the main results, while Section 3.2 extends them to a dynamic problem where allocations must be chosen over multiple periods.

3.1 Two static planning problems

A planner allocates input X_2 across a population indexed by X_1, choosing a possibly randomized rule—a coupling γ of X_1 and X_2 with the given marginal F_{X_1}—to maximize the expected valuation of the outcome,

$$\mathcal{W}(\gamma) = \iint q(x_1, x_2)\, \gamma(dx_1, dx_2).$$

Two constraints define the two problems. In the *fixed-marginal problem*, the distribution of X_2 is held at a target F_{X_2} and the planner only re-couples it to ability:

$$\max_{\gamma \in \Gamma(F_{X_1}, F_{X_2})} \mathcal{W}(\gamma), \tag{P1}$$

where $\Gamma(F_{X_1}, F_{X_2})$ is the set of couplings with marginals F_{X_1} and F_{X_2}. In the *budget problem*, only an aggregate resource constraint binds, at a non-decreasing unit cost function $b(x_2)$ and budget B:

$$\max_{\gamma \in \Gamma(F_{X_1})} \mathcal{W}(\gamma) \qquad \text{s.t.} \quad \int b(x_2)\, \gamma(dx_1, dx_2) \le B, \tag{P2}$$

where $\Gamma(F_{X_1})$ is the set of couplings whose X_1-marginal is F_{X_1}, the X_2-marginal being free. A coupling is *comonotone* if it places higher X_2 with higher X_1—equivalently $X_1 = F_{X_1}^{-1}(U)$, $X_2 = F_{X_2}^{-1}(U)$ for a common $U \sim \text{Uniform}(0,1)$—which reflects positive assortative matching of investment to ability.

Throughout, the technology is assumed productive, in the sense that more investment shifts the outcome distribution upward.

Assumption 1 (Monotone technology)**.** For every $y \in \mathcal{Y}$, the survival function $\overline{F}_{Y|X}(y \mid x_1, x_2)$ is non-decreasing in x_2: higher investment raises the outcome in the first-order stochastic dominance sense.

Under Assumption 1 the surplus of every non-decreasing valuation is non-decreasing in investment. The fixed-marginal result below does not use the assumption; the budget result does, since without it a supermodularity failure could hide behind investment levels that no planner facing non-decreasing costs would purchase. Two propositions then characterize assortative optimality: supermodularity of the surplus is necessary and sufficient in both problems.

Proposition 3 (Assortativity in the fixed-marginal problem)**.** *Let Q be non-decreasing and integrable. The surplus q is supermodular if and only if, for every pair of marginals (F_{X_1}, F_{X_2}), the comonotone coupling solves* (P1)*.*

This is the classical optimal-transport result, stated here without proof (Galichon, 2016, Th. 4.3): (P1) is an optimal transport problem on totally ordered supports, for which a supermodular surplus makes positive assortative matching optimal at any marginals, and optimality at every pair of marginals in turn forces supermodularity.

An interesting corollary of this proposition is that co-monotone couplings always yield better distributions of future output in a FOSD sense. To see why, fix $\overline{y} \in \mathcal{Y}$ and take $Q(y) = 1 - 1\{y \le \overline{y}\}$. From Proposition 3, $\mathbb{P}[Y \le \overline{y}]$ is minimized over all marginals that result from integrating the CDF F of Y given $X = (X_1, X_2)$ against the joint distribution $F_{X_1,X_2} \in \Gamma(F_{X_1}, F_{X_2})$. But since $\overline{y}$ was arbitrary, we get the following:

Corollary 1. *X_1 and X_2 are ordinal complements for Y if and only if for each pair of marginal CDFs F_{X_1} and F_{X_2}, the marginal of Y implied by a comonotone solution first-order stochastically dominates $\tilde{F}_Y(y) = \int_{\mathcal{X}_1 \times \mathcal{X}_2} F(y|x_1, x_2)\, dF_{X_1,X_2}(x_1, x_2)$ for any $F_{X_1,X_2} \in \Gamma(F_{X_1}, F_{X_2})$.*

The proof of this Corollary can be found in the Appendix. The following Proposition establishes that ordinal complementarity is necessary and sufficient for assortativity in the *budget problem*:

Proposition 4 (Assortativity in the budget problem)**.** *Let Q be non-decreasing and let Assumption 1 hold. The surplus q is supermodular if and only if, for every skill marginal F_{X_1}, every non-decreasing cost b, and every budget B, the budget problem* (P2) *admits a comonotone solution whenever it admits a solution.*

Sufficiency is straightforward because each pair (b, B) pins down a set of feasible marginal distributions over X_2, so that any comonotone coupling preserves the budget and weakly raises welfare by Proposition 3. Appendix A establishes necessity by showing that if q is not supermodular, one can find a triple (F_{X_1}, b, B) with a strictly counter-monotone solution.

Proposition 2 established an equivalence between ordinal complementarity and q being supermodular for all non-decreasing Q. Propositions 3 and 4 extend this equivalence class to a characterization of the solution properties of (P1) and (P2).

Corollary 2 (Ordinal complementarity and the planning problems)**.** *Under Assumption 1, the following are equivalent:*

(a) *X_1 and X_2 are ordinal complements in Y;*

(b) *for every non-decreasing bounded Q and every pair of marginals, the comonotone coupling solves the fixed-marginal problem* (P1)*;*

(c) *for every non-decreasing bounded Q, every skill marginal F_{X_1}, every non-decreasing cost b, and every budget B, the budget problem* (P2) *admits a comonotone solution whenever it admits a solution.*

The welfare conclusion is robust to the redistributive motive. Writing $Q(y) = c(y)^{1-\sigma}/(1-\sigma)$ with c a non-decreasing consumption schedule over outcomes, Q remains non-decreasing, so the assortative prescription stands no matter how large the curvature σ, even though channelling investment toward higher-ability children raises consumption inequality: under ordinal complementarity the production complementarity outweighs the redistributive motive.

3.2 Dynamic environments

In the previous subsection, we just showed that ordinal complementarity implies that at least two different types of Social Planner Problems have comonotone solution. Here, we show that comonotonicity of optimal policies under ordinal complementarity extends to dynamic contexts. For exposition we focus on the case of human capital investment in children. In this application, the inputs are current skill θ_t and investment I_t and the outcome is next-period skill θ_{t+1}, i.e. $(X_1, X_2, Y) = (\theta_t, I_t, \theta_{t+1})$. We prove that ordinal complementarity implies monotonicity of optimal investment allocations in a dynamic Social Planner Problem. A planner observes θ_t each period and chooses a (possibly random) allocation $F_{I_t|\theta_t}$ of investment to skill, subject to a per-period budget, to maximize the sum of valuations of skill along the path. The *sequential problem* is

$$\max_{\{F_{I_t|\theta_t}\}_{t=0}^{T}} \; E\left[\sum_{t=0}^{T} Q_t(\theta_t)\right] \quad \text{s.t.} \quad \theta_0 \sim F_{\theta_0}, \quad \theta_{t+1} \sim F_{t+1}(\cdot \mid \theta_t, I_t), \quad \int b(I_t)\, dF_{I_t} \leq B_t \;\forall t, \tag{6}$$

where each period's investment marginal F_{I_t} is the one induced by $F_{I_t|\theta_t}$ and the current skill distribution. The planner thus chooses, period by period, how to couple investment to the prevailing skill distribution, internalizing the effect on all future skill.

The monotonicity result for optimal solutions of this problem is given by the following proposition:

Proposition 5 (Dynamic comonotonicity). *Suppose each one-step technology $F_{t+1}(\cdot \mid \theta_t, I_t)$ is ordinally complementary and FOSD-monotone in (θ_t, I_t), and each Q_t is non-decreasing. Then, if a solution to the sequential problem (6) exists, it has a solution in which each period's coupling is comonotone, $I_t^* = F_{I_t}^{-1} \circ F_{\theta_t}(\theta_t)$.*

Proposition 5 is established in the following way: First, we prove an equivalence between the problem in 6 and its recursive formulation, akin to a Maximum Principle for dynamic programming Problems where the state is a marginal distribution (in this case of skills), and the control is a conditional distribution (in this case, the conditional distribution of investment given skills). Then, we establish monotonicity of the Value Function of the recursive formulation with respect to the state (which is a marginal distribution) in the first order stochastic dominance sense. Finally, we use Corollary 1 to show that if the recursive problem has a solution, then the comonotone coupling of the marginal induced by said solution has to be a solution too, given that it yields a better distribution of skills tomorrow, which by monotonicity implies a higher value tomorrow, and hence a higher value (since the flow payoff only depends on the contemporaneous distribution).

A second dynamic problem—*optimal tracking*, in which the planner observes only initial skill and assigns each child a whole sequence of investments—also admits a monotone solution, but the argument requires a strict form of ordinal complementarity and a correspondingly strict form of the equivalence in Proposition 2. Online Appendix E states the problem, develops the strict machinery, and proves the result.

Note that here we have only established sufficiency of ordinal complementarity for monotonicity of solutions to dynamic social planner problems. Necessity is trivial to establish by letting $T = 1$ and using necessity in the static contexts.

4 Testing ordinal complementarity

This section turns the properties of Section 2 into tests, for two discrete inputs X_1, X_2 and a discrete outcome Y, initially following the moment-inequality literature (Andrews and Soares, 2010; Romano et al., 2014). Ordinal complementarity (Definition 1) is a finite system of linear inequalities and can be tested directly (Section 4.1). The weaker properties of Section 2.2—non-substitutability and convex complementarity—are characterized through the difference-in-differences of a valuation (Section 4.2) and tested by a linear program and a second moment-inequality system respectively (Section 4.2). Finally, Section 4.3 briefly describes an alternative to the moment inequality tests based on the projection method outlined in Fang and Seo (2021).

4.1 Testing ordinal complementarity

Take supports $\mathcal{X}_1 = \{1, \dots, r_1\}$, $\mathcal{X}_2 = \{1, \dots, r_2\}$, and $\mathcal{Y} = \{1, \dots, r_Y\}$, and stack the conditional probabilities $\pi(j, k, \ell) = \mathbb{P}(Y = j \mid X_1 = k, X_2 = \ell)$ into a vector $\Pi \in \mathbb{R}^p$; thus Π holds the transition *masses*. Dropping one redundant probability per (k, ℓ) cell leaves $p = (r_Y - 1)\, r_1 r_2$ free parameters. Write $\overline{F}_{Y|X}(j \mid k, \ell) = \mathbb{P}(Y > j \mid X_1 = k, X_2 = \ell) = \sum_{j' > j} \pi(j', k, \ell)$ for the conditional survival function and, for a cutpoint j and adjacent cell (k, ℓ), define:

$$\begin{aligned} \Delta^2 \overline{F}_{Y|X}(j \mid k, \ell) = \overline{F}_{Y|X}(j \mid k{+}1, \ell{+}1) + \overline{F}_{Y|X}(j \mid k, \ell) \\ - \overline{F}_{Y|X}(j \mid k{+}1, \ell) - \overline{F}_{Y|X}(j \mid k, \ell{+}1) \end{aligned} \tag{7}$$

for its mixed second cross-difference in (X_1, X_2). By Definition 1, X_1 and X_2 are ordinal complements if and only if $\overline{F}_{Y|X}(j \mid \cdot)$ is supermodular at every cutpoint, i.e. $\Delta^2 \overline{F}_{Y|X}(j \mid k, \ell) \geq 0$ for all j, k, ℓ. Supermodularity is characterized by these *adjacent* cross-differences alone; wider rectangles are non-negative sums of them. Collecting the adjacent conditions into $\Delta = S\Pi \in \mathbb{R}^d$ for a known $d \times p$ matrix S with entries in $\{-1, 0, 1\}$—there are $d =$

$(r_Y - 1)(r_1 - 1)(r_2 - 1)$ of them (e.g. 36 when $r_Y = r_1 = 4$, $r_2 = 5$, against 180 for the full set of all rectangle pairs)—the null hypothesis is

$$H_0: \ \Delta = S\Pi \geq 0_{d\times 1}.$$

4.1.1 Test statistic and critical value.

Let $\hat{\Pi}$ be an asymptotically normal estimator with $\sqrt{N}(\hat{\Pi} - \Pi) \rightsquigarrow \mathcal{N}(0, \Sigma_\Pi)$ and a consistent $\hat{\Sigma}$; Section 5 obtains $\hat{\Pi}$ and $\hat{\Sigma}$ from the score of the observed-data likelihood. Then $\hat{\Delta} = S\hat{\Pi}$ satisfies $\sqrt{N}(\hat{\Delta} - \Delta) \rightsquigarrow \mathcal{N}(0, S\Sigma_\Pi S')$. Let D_Π be diagonal with $[D_\Pi]_{jj} = [S\Sigma_\Pi S']_{jj}^{1/2}$ and $\Omega_\Pi = D_\Pi^{-1} S\Sigma_\Pi S' D_\Pi^{-1}$ the implied correlation matrix, with sample analogues $\hat{D}, \hat{\Omega}$. Because a violation of H_0 is a *negative* entry of $\hat{\Delta}$, the modified-method-of-moments statistic penalizes the negative part,

$$\hat{T}_{\mathrm{c}} = \left\| \left(\sqrt{N}\, \hat{D}^{-1}\hat{\Delta} \right) \wedge 0_{d\times 1} \right\|^2 = \sum_{k=1}^{d} \min\Big(0,\ \sqrt{N}[\hat{D}^{-1}\hat{\Delta}]_k\Big)^2,$$

and the test rejects when $\hat{T}_{\mathrm{c}} > c_\alpha(\hat{\Omega})$, where $c_\alpha(\Omega)$ is the $(1-\alpha)$-quantile of $\|Z \wedge 0\|^2$ for $Z \sim \mathcal{N}(0, \Omega)$. The critical value is evaluated at the least-favourable point of H_0 (all constraints binding, $\Delta = 0$). Standard arguments (Andrews and Guggenberger, 2009) give pointwise and uniform asymptotic size control,

$$\limsup_{N\to\infty} \sup_{\Pi:\, S\Pi \geq 0} E_\Pi[\phi(\hat{T}_{\mathrm{c}}, \hat{\Omega})] \leq \alpha.$$

The symmetric statistic $\hat{T}_{\mathrm{s}}$ that penalizes the *positive* part of $\hat{\Delta}$ tests ordinal *substitutability* (H_0: $S\Pi \leq 0$, every survival cross-difference non-positive); by symmetry of the least-favourable null the two statistics share the critical value. Both are reported in Section 6.

4.1.2 Moment selection.

Because the least-favourable simulation is conservative when only a subset of inequalities are close to binding, a moment-selection refinement improves power. For small $\beta < \alpha$, retain the standardized moments that remain on the violated (negative) side after accounting for sampling uncertainty,

$$\hat{\mu}_\beta = \left(\sqrt{N}\, \hat{D}^{-1}\hat{\Delta} + \kappa_\beta(\hat{\Omega}) \right) \wedge 0_{d\times 1},$$

where $\kappa_\beta(\Omega)$ is a small positive slack, and compare $\hat{T}_c$ to the $(1-\alpha+\beta)$-quantile of $\|Z \wedge 0\|^2$ with $Z \sim \mathcal{N}(\hat{\mu}_\beta, \hat{\Omega})$. The refinement retains asymptotic size control (Andrews and Soares, 2010; Romano et al., 2014) and is the procedure used in the application; the recommended tuning is $\beta = \alpha/10$.

4.2 Testing the weaker properties

The weaker properties of Section 2.2 are statements about $E[Q(Y) \mid x_1, x_2]$ for restricted classes of Q. This section rewrites this condition in terms of second differences and uses it to develop testable characterizations of ordinal non-substitutability and convex ordinal complementarity (the two weakenings of ordinal complementarity introduced by Section 2.2. When Y and X have finite support, these turn out to be useful and relatively straightforward.

For an increasing valuation, write the increments of Q as $\Delta Q(k) = Q(k+1) - Q(k) \geq 0$. Summation by parts gives $q(x_1, x_2) = Q(1) + \sum_k \Delta Q(k)\, \overline{F}_{Y|X}(k \mid x_1, x_2)$, so the second differences in $\mathbb{E}[Q(Y) \mid x_1, x_2]$ is a ΔQ-weighted sum of the survivor function second differences:

$$\Delta^2 q(x_1, x_2) = \sum_k \Delta Q(k)\, \Delta^2 \overline{F}_{Y|X}(k \mid x_1, x_2). \tag{8}$$

The expectation is supermodular for a given Q iff (8) is ≥ 0 at every (x_1, x_2), submodular iff ≤ 0. Each of the three properties can be viewed through this equation. (a) If *every* cutpoint difference-in-difference is non-negative the sum is non-negative for every increasing Q—this is ordinal complementarity again. (b) Asking only that *some* increasing Q make the sum non-negative on every cell is an existence question, linear in ΔQ, hence a linear program. (c) Restricting Q to convex increments—non-decreasing $\Delta Q(k)$—re-weights the cutpoints toward the top of the scale, and summation by parts turns the condition into one on the cumulative sums of the cross-differences.

Since (8) is invariant to the scale of ΔQ, no generality is lost in normalizing the increments to the simplex $\mathcal{Q} = \{\Delta Q \geq 0 : \mathbf{1}'\Delta Q = 1\}$. The existence question (b) can then be chracterized by the pair of linear programs

$$t^+(F) = \max_{\Delta Q \in \mathcal{Q}} \min_{(x_1, x_2)} \Delta^2 q(x_1, x_2), \qquad t^-(F) = \min_{\Delta Q \in \mathcal{Q}} \max_{(x_1, x_2)} \Delta^2 q(x_1, x_2), \tag{9}$$

and the restricted question (c) is a system of inequalities on cumulative sums. The next proposition formalizes a condition for testing.

Proposition 6 (Non-substitutability and rationalizability). *(i) There exists an increasing non-constant Q whose expectation is supermodular if and only if $t^+(F) \geq 0$; the maximizing ΔQ^* is such a valuation. Symmetrically, some increasing non-constant Q has submodular surplus iff $t^-(F) \leq 0$.*

(ii) Hence the technology is ordinally non-substitutable *(Definition 2) iff $t^-(F) > 0$, and* ordinally non-complementary *iff $t^+(F) < 0$.*

This characterization suggests a natural test by bootstrapping the distribution of t^+ and t^- and ruling on the mutually exclusive partitions of the space of distributions based on the distribution of the test statistic.

The proposition below rearranges (8) to derive a set of inequalities that characterize convex ordinal complementarity and concave ordinal substitutability.

Proposition 7 (Convex complementarity and concave substitutability). *Let $T_j(x_1, x_2) = \sum_{k \geq j} \Delta^2 \overline{F}_{Y|X}(k \mid x_1, x_2)$ be the suffix sums of the cutpoint cross-differences and $L_j(x_1, x_2) = \sum_{k \leq j} \Delta^2 \overline{F}_{Y|X}(k \mid x_1, x_2)$ the prefix sums.*

(i) X_1, X_2 are convex-ordinally complementary in Y (Definition 3) if and only if $T_j(x_1, x_2) \geq 0$ for every cutpoint j and adjacent cell (x_1, x_2).

(ii) X_1, X_2 are concave-ordinally substitutable if and only if $L_j(x_1, x_2) \leq 0$ for every j and cell.

Crucially, both T_j and L_j are *linear* in Π—they are partial sums of the entries of $S\Pi$—so Proposition 7 is a moment-inequality null of exactly the form tested in Section 4.1, with S replaced by the suffix- (or prefix-) summed constraint matrix. Convex complementarity is therefore tested by the same moment inequality approach.

4.3 A projection test

Every one-sided null in this section—ordinal complementarity and substitutability (Section 4.1) and their convex and concave refinements (Proposition 7)—is a system of linear inequalities on the survival probabilities, confining the survivor vector to a *convex cone*

$$\Lambda = \{ s \in \mathbb{R}^p : \ As \geq 0 \},$$

where s stacks the $\overline{F}_{Y|X}(j \mid k, \ell)$ and A is the relevant constraint matrix. The structure of this space lends itself to a projection-based test, due to Fang and Seo (2021), which the results below report alongside the moment-inequality tests. Intuitively, the test rejects the null when the minimum Euclidean distance from the estimate $\hat{s}$ to the cone (the projection) is sufficiently large and unlikely to be caused by sampling error. Formally, the test statistic is:

$$\hat{T}_{\mathrm{FS}} \;=\; \sqrt{N}\,\left\|\hat{s} - \Pi_\Lambda \hat{s}\right\|_2, \qquad \Pi_\Lambda \hat{s} \;=\; \operatorname*{arg\,min}_{h:\, Ah \geq 0} \|\hat{s} - h\|_2,$$

where Π_Λ is the projection operator that maps vectors to their closest point in the cone. As in the moment inequality test, the most conservative approach would be to test against the least favorable null (the vertex of the cone, where every inequality binds). This is needlessly conservative. Fang and Seo (2021) provide a data-driven approach that bootstraps the statistic around a point that shrinks the estimated location towards the vertex at an appropriate rate.

Let $\hat{\mathbb{G}}_b = \sqrt{N}(\hat{s}^*_b - \hat{s})$ be a single draw from the re-centered bootstrap distribution. The critical value for the test statistic is the $(1-\alpha)$-quantile from the bootstrapped distribution of statistics:

$$\hat{T}^*_b \;=\; \left\|\Big(\hat{\mathbb{G}}_b + \hat{\kappa}_n \Pi_\Lambda \hat{s}\Big) - \Pi_\Lambda\Big(\hat{\mathbb{G}}_b + \hat{\kappa}_n \Pi_\Lambda \hat{s}\Big)\right\|_2, \qquad b = 1, \ldots, B.$$

The tuning parameter $\hat{\kappa}_n \in [0, \sqrt{N}]$ is the key to the test. Setting $\hat{\kappa}_n = 0$ recovers the distribution at the least-favorable vertex, while larger values let the estimated location absorb inequalities that are far from binding. This is the same role that moment selection plays in Section 4.1. The convex cone structure ensures shrinking toward the vertex can only raise the critical value, so an imperfect $\hat{\kappa}_n$ errs toward conservatism rather than size distortion, and the test is uniformly valid yet asymptotically nonconservative as long as $\kappa_n/\sqrt{N} \to 0$ (Fang and Seo, 2021, Theorems 3.1–3.2).

To choose the tuning parameter, Fang and Seo suggest a data-driven choice $\hat{\kappa}_n = \sqrt{N}\, c_n / \hat{\tau}_{1-\gamma_n}$, with $\hat{\tau}_{1-\gamma_n}$ the $(1-\gamma_n)$-quantile of the noise magnitudes $\|\hat{\mathbb{G}}_b\|_2$: the rate of convergence $\sqrt{N}$ deflated by an estimate of the noise. The tests below set $\gamma_n = 0.01/\log N$, as their simulations recommend, and verify that every reported verdict is unchanged for $\gamma_n \in \{1/N,\, 0.01,\, 0.1\}$; $c_n = 1$ is appropriate in regular $\sqrt{N}$-consistent settings such as this one.

5 Identification and estimation

The empirical object of interest is the technology of skill formation $\mathbb{P}(\theta_{t+1} \mid \theta_t, I_t, \dots)$, where both skills and investment are latent and discrete and are observed only through noisy measures. This section lays out the dynamic model along with the strategy for identification and estimation. The identification argument follows Garcia-Vazquez (2021); the measurement structure follows Cunha et al. (2010) and the nonparametric mixture-identification of Hu and Schennach (2008) and Bonhomme and Robin (2009).

5.1 Model

Identification (Section 5.2) is nonparametric. For stable estimation and a manageable state space, the application imposes parametric structure on transitions, choices, and emissions. The time-varying latent variables are (θ_t, I_t) for $t = 1, \dots, 7$. Three time-invariant latent variables enter the model: parental cognitive skill θ^P (measured by six ASVAB sections), child heterogeneity ω (no own measures), and an income type ε_Y. A configuration of these three—written $z = (\theta^P, \omega, \varepsilon_Y)$—is a *type*; the model is specified conditional on the type. The inclusion of the latent type ω follows Cunha et al. (2010) to allow for potentially endogenous investment choices.

5.1.1 Technology

The transition $\mathbb{P}(\theta_{t+1} \mid \theta_t, I_t, \theta^P, \omega)$ is the object of interest, and its functional form is chosen so that the estimated complementarity is not arbitrarily or accidentally imposed. A one-hidden-layer network with K probit units forms a flexible index in the inputs,

$$\mu(\theta, I, \theta^P, \omega) = \sum_{k=1}^{K} w_k \, \Phi\Big(\alpha_{1k}\theta + \alpha_{2k} I + \alpha_{3k}\theta^P + \alpha_{4k}\omega + b_k\Big), \qquad w_k, \alpha_{2k} > 0,$$

and the survival function passes this index through a *robit* link—the cumulative distribution function T_ν of a Student-t with ν degrees of freedom, used in place of the normal of a probit (hence "robit," for robust probit):

$$\overline{F}_{Y|X}(j \mid \theta, I, \theta^P, \omega) = T_\nu\Big(c_j + \mu(\theta, I, \theta^P, \omega)\Big), \qquad c_1 > \cdots > c_{r_\theta - 1}.$$

The fat tails of the Student-t (here $\nu = 7$) are the point of the choice. A fat-tailed link is close to linear over most of its range and bends toward 0 or 1 only far out in the tails, so the survival probabilities track the index $c_j + \mu$ almost linearly except when they are very close to 0 or 1. The monotonicity and complementarity structure of the flexible index is thereby inherited by the probabilities with little distortion from the link's own curvature; a thin-tailed link such as the probit or logit would impose an S-shaped bend over the whole range that could itself create or mask complementarity in the estimated technology. The production technology (transitions) is grouped into early childhood (ages 0–5) and school age (ages 6–13). Parental skill θ^P and heterogeneity ω enter each unit's index.

5.1.2 Investment

Investment is a behavioural choice, governed by a conditional choice probability (CCP) that is itself an ordered probit. Permanent log income is a linear function of the type, and observed income adds transitory noise:

$$\log y_t = \underbrace{\beta_0 + \beta_P \theta^P + \beta_\omega \omega + \delta\, \varepsilon_Y}_{\text{permanent log income}} + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma^2) \text{ iid given the type,}$$

where u_t is measurement error—or transitory deviation from permanent income—that does not affect the investment decision. Investment depends on permanent income, hence on $(\theta^P, \omega, \varepsilon_Y)$. Because parental skill and child heterogeneity already enter the CCP in their own right, only the income type need be added to capture the income channel: the index for the ordered probit defining choice probabilities is

$$\beta_1 \theta_t + \beta_3 \theta^P + \beta_4 \omega + \beta_5\, \varepsilon_Y.$$

The strength of the income channel is the ratio β_5/δ: a one-unit change in the income type shifts the CCP index by β_5 and permanent log income by δ, so β_5/δ is the semi-elasticity of the investment index to permanent log income.

5.1.3 Measurement

Every noisy measure of skill or investment follows an ordered probit, $\mathbb{P}(\tilde{x}^m = k \mid \text{state} = j) = \Phi(\tau_{m,k} - \alpha_m j)$, increasing in the latent state j whenever the loading $\alpha_m > 0$. Parental skill θ^P is measured under the same ordered-probit form; observed log income is a noisy measure of the type through the income equation above; and the four adult outcomes—high-school completion, college completion, earnings quintile, and employment at ages 25–30—are modelled as $\mathbb{P}(Z \mid \theta_T, \omega)$, serving both as the anchors Q for the policy experiments and as the only signals of the latent type ω, which carries no direct measures of its own. The benchmark cardinalities are $r_\theta = 4$, $r_I = 5$, $r_P = r_\omega = 3$, $\varepsilon_Y \in \{1, 2, 3\}$, and $K = 4$ hidden units, the last selected by BIC over $K \in \{3, 4, 5\}$.

5.2 Identification

Children are observed for $T \geq 2$ developmental periods. Each period yields several discrete noisy measures of cognitive skill $\tilde{\theta}_t^m$ and of investment $\tilde{I}_t^m$. The measures are *pure*: conditional on the true latent state, each is independent of all other measures, of other periods, and of all other latent variables. The emission matrices $(P_I^m)_{ij} = \mathbb{P}(\tilde{I}^m = i \mid I = j)$ and $(P_\theta^m)_{ij} = \mathbb{P}(\tilde{\theta}^m = i \mid \theta = j)$ have full column rank, and at least one measure of each kind is monotone, in the sense that the probability of its highest category is increasing in the latent state.

The identification argument maps the pair (θ, I) to a single index $S_t = G^S(\theta, I) = (\theta - 1)r_I + I$, and pairs the period-$t$ measures into composite measures $Y^m = G^m(\tilde{\theta}^m, \tilde{I}^m)$. Purity makes Y^1, Y^2, Y^3 conditionally independent given S_t, and the composite emission matrix factors as $P^m = P_\theta^m \otimes P_I^m$, which inherits full column rank. Given full support of the latent state at each period, Theorem 1 of Garcia-Vazquez (2021) identifies the marginal $\mathbb{P}(\theta_t, I_t)$ and the joint transition kernel K_t, from which $\mathbb{P}(\theta_{t+1} \mid \theta_t, I_t) = \sum_{I_{t+1}} K_t$ follows; the monotone measure fixes the labelling of categories. Because the technology (the transition function) is left unrestricted, the discrete relabelling at each age does not run afoul of the critique of Agostinelli and Wiswall (2016). The argument accommodates the additional discrete latents—parental skill, child heterogeneity, the income type—by enlarging the composite index, provided each carries (or, for heterogeneity revealed through the panel and adult outcomes, is pinned down by) enough conditionally independent signals.

5.3 Estimation

Estimation uses expectation–maximization. The two time-invariant skill type (θ^P, ω) and the income type ε_Y are handled by a mixture-over-types E-step: for each value of the combined type $z = (\theta^P, \omega, \varepsilon_Y)$, a forward–backward (Baum–Welch) pass returns the posteriors over all latent states. The M-step updates the parameters defining the technology, household investment policies (the CCPs), and the emission probabilities. To strengthen the relatively weak identification of the large number of parameters in the technology, the preferred specification imposes a small ridge penalty on the weights in the hidden layer of the neural network. A *child bootstrap* with 200 re-samples quantifies the sampling variation of the estimates and is, in practice, more conservative than the corresponding asymptotic approximation.

6 Empirical illustration

6.1 Data

The data come from the NLSY-79 Child and Young Adult supplement, a biennial panel of children of the NLSY-79 women observed between 1986 and 2016, with roughly 11,500 children. The panel is organized into seven developmental periods spanning ages 0–13 in two-year bins. Each child-period carries cognitive measures (the Motor and Social Development scale at young ages; PPVT, PIAT, and Digit Span at older ages), investment measures (the HOME-SF inventory, Parts A–D), and—cross-sectionally—six ASVAB sections measuring parental skill. Household income (log, biennial) and four adult outcomes (high-school completion, college completion, earnings quintile at ages 25–30, and employment at ages 25–30) complete the data. Continuous scores are discretized by pooled quantile cut-offs across ages, and every period carries at least three noisy measures, as identification requires. Table 1 summarizes coverage.

Period	Ages	Skill sources	Investment
1	0–1	MSD	HOME Part A
2	2–3	MSD, PPVT	HOME Part A/B
3	4–5	MSD, PPVT, PIAT (Math/Recog/Comp)	HOME Part B
4	6–7	PPVT, PIAT, Digit Span (F/B)	HOME Part C
5	8–9	PPVT, PIAT, Digit Span	HOME Part C
6	10–11	PPVT, PIAT, Digit Span	HOME Part C/D
7	12–13	PPVT, PIAT, Digit Span	HOME Part D

Table 1: Measurement coverage by developmental period.

6.2 The estimated technology

Figure 2 plots the estimated, type-marginalized survival functions $\overline{F}_{Y|X}(k \mid \theta, I)$ with confidence bands from a child-level bootstrap. The technology is monotone in both skill and investment, but the spacing of the survival curves across investment levels is far from uniform across the skill grid, a sign of non-constant complementarity. Ordinal complementarity would produce increasing differences in these survival probabilities, and one can informally visually reject ordinal complementarity based on Figure 2. The strong ceiling effects for some cut-offs make complementarity very hard to obtain: a echoing a reminiscent finding in Heckman et al. (2026).

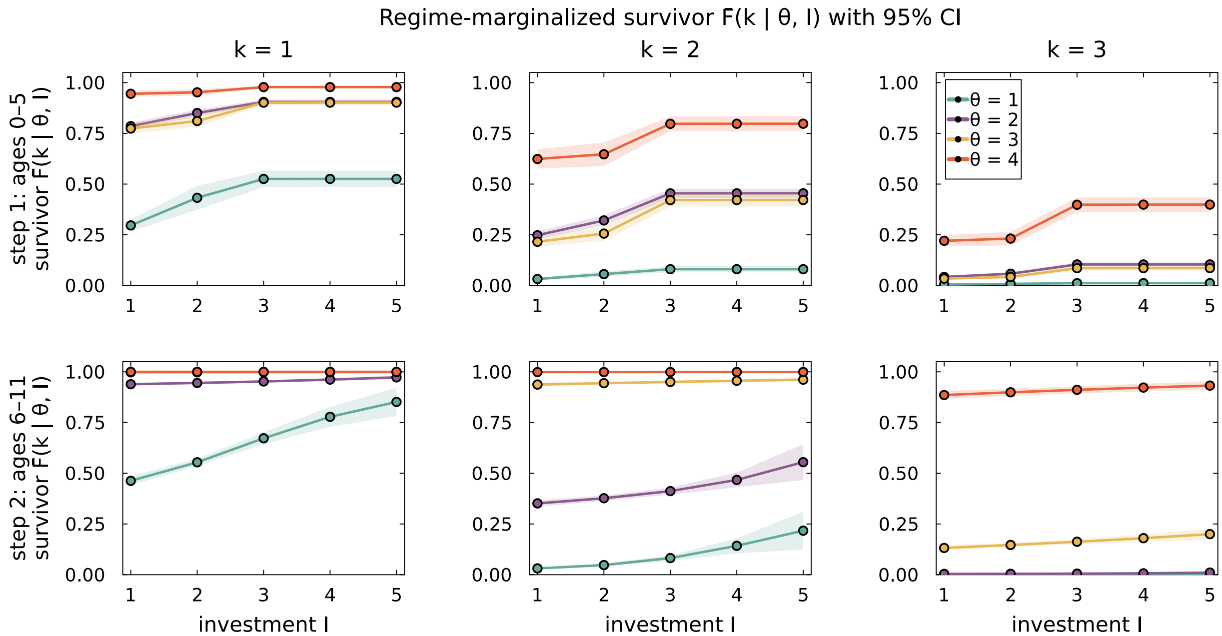

Figure 2: Estimated type-marginalized survival functions $\overline{F}_{Y|X}(k \mid \theta, I)$ by transition group (rows) and skill threshold k (columns). Shaded bands are 95% child-bootstrap confidence intervals ($B = 200$, from the regularized fit). Colours index current skill θ.

6.3 Does ordinal complementarity hold?

Table 2 reports the results of the three tests of ordinal complementarity: the least favorable moment inequality test, the version with data-dependent moments due to Romano et al. (2014), and the projection test of Fang and Seo (2021).

While the tests do not completely agree on their verdicts, the overall patterns of results and p-values suggests that neither ordinal complementarity nor ordinal substitutability holds for younger children. The data also reject the possibility of ordinal complementarity for older (school-aged) children. However none of the tests can rule out the possibility of ordinal substitutability for older children. This pattern of results will eventually be borne out by the shape of optimal policies in Section 6.7.

A *local* reading provides some exploration of the inconsistent structure of production

Transition group	H_0	MMM (least favorable) $\hat{T}_{\text{MMM}}$	$c^{\text{LF}}_{0.05}$	rej.	MMM (RSW) $c^{\text{RSW}}_{0.05}$	rej.	Fang–Seo projection $\hat{T}_{\text{FS}}$	$\hat{c}_{0.05}$ (p)	rej.
ages 0–5	complementarity	24.0	43.3	no	43.7	no	12.6	12.4 (0.04)	yes
	substitutability	51.5	43.3	yes	45.0	yes	18.9	12.7 (0.01)	yes
ages 6–13	complementarity	99.9	38.4	yes	39.4	yes	30.2	36.9 (0.07)	no
	substitutability	20.8	38.4	no	39.4	no	4.4	18.8 (0.93)	no

Table 2: Ordinal complementarity (H_0: supermodularity in distribution) and substitutability (H_0: submodularity) of the technology $\overline{\Pi}(\theta' \mid \theta, I)$, by transition group: least-favorable, moment-selection, and projection critical values. $\hat{T}_{\text{MMM}}$ is the modified-method-of-moments statistic of Section 4.1 on the 36 adjacent survival cross-differences (child-bootstrap covariance, $B = 200$), with critical values $c^{\text{LF}}_{0.05}$ simulated from the least-favorable null and $c^{\text{RSW}}_{0.05}$ from the moment-selection refinement of Romano et al. (2014) with first-stage size $\beta = \alpha/10$. $\hat{T}_{\text{FS}}$ is the projection statistic of Fang and Seo (2021) described in Section 4.3, with critical value $\hat{c}_{0.05}$ and bootstrap p-value computed from the same $B = 200$ child-bootstrap draws and tuning $\gamma_n = 0.01/\log N$, $c_n = 1$ (verdicts are unchanged for $\gamma_n \in \{1/N, 0.01, 0.1\}$). Rejection is evidence against the named property.

within each age group, classifying each (θ_t, I_t) increment by the sign of

$$\overline{F}_{Y|X}(\cdot \mid \theta+1, I+1) - \overline{F}_{Y|X}(\cdot \mid \theta+1, I) - \overline{F}_{Y|X}(\cdot \mid \theta, I+1) + \overline{F}_{Y|X}(\cdot \mid \theta, I).$$

Figure 3 maps the local classification over the skill–investment grid for each group: suggesting that regions of local complementarity and local substitutability coexist within the same estimated technology.

6.4 Can any valuation rationalize assortative investment?

Table 2 asks whether *every* increasing valuation makes skill and investment complementary. Proposition 6 asks the weaker existence questions—whether *some* increasing valuation results in a supermodular expectation, or some rationalizes a submodular one—by solving the linear programs (9) on the estimated $\hat{\overline{\Pi}}$ and bootstrapping their values. Table 3 reports the result.

Three conclusions follow. First, neither stage is comonotone-rationalizable: $\hat{t}^+ < 0$ in both, making the technology ordinally non-complementary. This is a strong rejection of ordinal complementarity that supplements the results in 2.

Second, early childhood (ages 0–5) is *ordinally non-substitutable* (Definition 2): the bootstrapped test statistic is strictly positive ($\hat{t}^- = +0.019$) and stays positive in every bootstrap resample ($\Pr(t^- \leq 0) = 0$), so no nondecreasing Q can produce a submodular expectation.

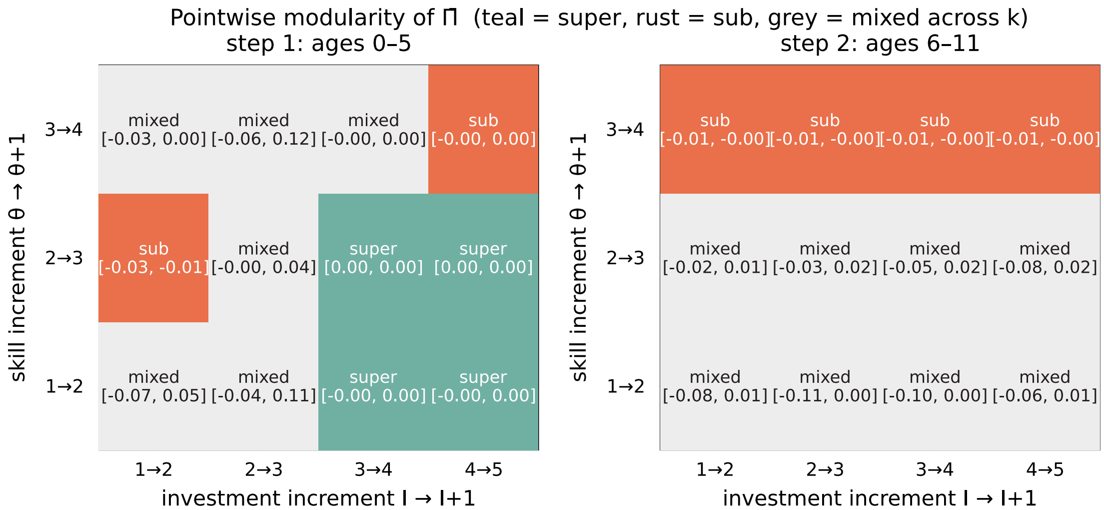

Figure 3: Pointwise difference-in-difference sign over (θ_t, I_t) increments, by transition group. Each cell is classified as locally super-, sub-, or mixed-modular, with the minimum and maximum DID across cutpoints reported.

Transition group	Rationalizable as	$\hat{t}^*$	95% CI	Pr	verdict
ages 0–5	comonotone ($t^+ \geq 0$)	-0.008	$[-0.017, -0.000]$	0.03	no
	counter-monotone ($t^- \leq 0$)	$+0.019$	$[+0.001, +0.069]$	0.00	no
ages 6–13	comonotone ($t^+ \geq 0$)	-0.008	$[-0.068, -0.003]$	0.00	no
	counter-monotone ($t^- \leq 0$)	-0.004	$[-0.007, +0.014]$	0.85	yes

Table 3: Rationalizability of assortative and anti-assortative allocations (Proposition 6). t^+ (t^-) is the LP value (9) whose sign decides whether some increasing valuation makes the optimal coupling comonotone (counter-monotone); $\hat{t}^*$, its 95% percentile interval from the child-level bootstrap ($B =$ 200 resamples of $\overline{\Pi}$), and the bootstrap probability Pr of the favourable sign ($\Pr(t^+ \geq 0)$ or $\Pr(t^- \leq 0)$) are unrestricted over the valuation shape.

Together with the first conclusion this means that optimal allocations in the two planner's problems are unlikely to be monotonic, given the evidence that expectations can be neither super- nor submodular.

Third, school age (ages 6–13) is the one stage where counter-monotone investment may be guaranteed: the test statistic is non-positive ($\hat{t}^- = -0.004$), in concordance with the evidence presented in Table 2.

6.5 Complementarity in the income-induced transition

The tests so far concern the *technology* $\Pi(\theta_{t+1} \mid \theta_t, I_t)$, whose second argument is latent investment. Replacing investment with *exogenous household income* gives a policy-relevant counterpart: composing the estimated technology with the investment policy yields the reduced-form transition a child faces when income is set exogenously,

$$\Pi^{\text{ind}}(\theta_{t+1} \mid \theta_t, y) = \sum_{\theta^P,\omega} \pi_{P\omega}(\theta^P,\omega) \sum_{I_t} \mathbb{P}\Big(I_t \mid \theta_t, \theta^P, \omega, \varepsilon_Y^{\text{eff}}(y, \theta^P, \omega)\Big)\, \Pi(\theta_{t+1} \mid \theta_t, I_t, \theta^P, \omega). \tag{10}$$

Because income enters investment only through the type ε_Y (Section 5.1), setting income to y means setting ε_Y to the value that reproduces y in the child's type, $\varepsilon_Y^{\text{eff}}(y, \theta^P, \omega) = (\log y - \beta_0 - \beta_P\theta^P - \beta_\omega\omega)/\delta$ — the same inversion used in the income-transfer counterfactual of Section 6.7. The type (θ^P, ω) is integrated out with its *unconditional* marginal $\pi_{P\omega}$: income and current skill are set exogenously, so the type is independent of them.[9] Income is discretized to five model-implied quantiles (\$9.9k–\$139k); the result is structurally a transition with income in place of investment, so the tests of Sections 4 and 4.2 apply verbatim. Unlike the technology test, this object folds in how investment responds to resources.

		MMM (least favorable)			MMM (RSW)		Fang–Seo projection		
Transition group	H_0	$\hat{T}_{\text{MMM}}$	$c_{0.05}^{\text{LF}}$	rej.	$c_{0.05}^{\text{RSW}}$	rej.	$\hat{T}_{\text{FS}}$	$\hat{c}_{0.05}$ (p)	rej.
ages 0–5	complementarity	168.2	44.1	yes	36.4	yes	10.8	4.8 (0.00)	yes
	substitutability	365.2	44.1	yes	43.0	yes	14.0	6.9 (0.01)	yes
ages 6–13	complementarity	127.3	43.2	yes	44.7	yes	26.3	22.2 (0.04)	yes
	substitutability	32.1	43.2	no	44.0	no	4.0	14.1 (0.73)	no

Table 4: Ordinal complementarity and substitutability of the income-induced transition $\Pi^{\text{ind}}(\theta' \mid \theta, y)$ in (10), by transition group: same statistics, constraint set, critical values, and tuning as Table 2, with income in place of investment.

Two features stand out (Table 4). First, the developmental contrast of Table 2 replicates. The complementarity null is rejected jointly in both stages—income substitutes for skill somewhere—but the substitutability null is rejected only in early childhood; at school age it is *not* rejected ($\hat{T}_{\text{s}} = 32 < 43$), so the school-age income-induced transition is substitution-dominant, exactly as the technology is. The statistics are an order of magnitude larger than

[9] Weighting instead by the skill-conditional $\pi_{P\omega}(\cdot \mid \theta_t)$ — the type mix among children observed at skill θ_t — would reintroduce selection on permanent types and answer a descriptive rather than a structural question.

in Table 2 because the policy channel amplifies the latent interaction: lower-skill children's investment responds more strongly to income, concentrating significant income-substitution at the lower thresholds (eleven of twelve adjacent cells at $\theta' \geq 2$ at school age). Unlike the technology, the induced verdicts are unanimous across critical values: Table 4 shows the least-favorable, moment-selection, and Fang–Seo projection tests all reject complementarity in both stages and substitutability in early childhood only.

Second, the rationalizability tests (9), run on Π^{ind}, reinforce these implications (Table 5). Early childhood is again ordinally *non-substitutable* ($\hat{t}^- = +0.002$, $\Pr(t^- \leq 0) = 0.01$): no monotone valuation rationalizes anti-assortative income. School age is the opposite: the counter-monotone value is non-positive in every bootstrap resample ($\hat{t}^- = -0.003$, $\Pr(t^- \leq 0) = 1.00$).

Transition group	Rationalizable as	$\hat{t}^*$	95% CI	Pr	verdict
ages 0–5	comonotone ($t^+ \geq 0$)	-0.002	$[-0.004, +0.004]$	0.09	no
	counter-monotone ($t^- \leq 0$)	$+0.002$	$[+0.001, +0.010]$	0.01	no
ages 6–13	comonotone ($t^+ \geq 0$)	-0.010	$[-0.049, -0.005]$	0.00	no
	counter-monotone ($t^- \leq 0$)	-0.003	$[-0.004, -0.001]$	1.00	yes

Table 5: Rationalizability of assortative and anti-assortative *income* allocations on the income-induced transition $\Pi^{\text{ind}}(\theta' \mid \theta, y)$ of (10), the counterpart of Table 3 with income in place of investment. t^+ (t^-) is the linear-program value (9) whose sign decides whether some increasing valuation makes the optimal income coupling comonotone (counter-monotone); $\hat{t}^*$, its 95% percentile interval from the child-level bootstrap, and Pr of the favourable sign ($\Pr(t^+ \geq 0)$ or $\Pr(t^- \leq 0)$) are unrestricted over the valuation shape. Early childhood is ordinally non-substitutable ($t^- > 0$); school age is decisively counter-monotone-rationalizable ($\Pr(t^- \leq 0) = 1.00$, against the boundary 0.85 for the technology in Table 3), so a concave anchor rationalizes progressive income transfers (an existence statement, not a claim about *every* concave anchor).

6.6 Convex complementarity and concave substitutability

The rationalizability programs ask whether *some* valuation can hold the inputs as complements; the shape-refined properties of Section 2.2 ask whether *every* convex (or concave) valuation does. By Proposition 7 these are moment-inequality nulls—convex complementarity is $T_j(\theta, I) = \sum_{k \geq j} \Delta^2 \overline{F}_{Y|X}(k \mid \theta, I) \geq 0$ on every cell and cutpoint, concave substitutability is $L_j = \sum_{k \leq j} \Delta^2 \overline{F}_{Y|X}(k \mid \theta, I) \leq 0$—and are tested by the statistic of Section 4.1 applied to the cumulative cross-differences, alongside the projection test of Section 4.3, which applies verbatim to the convex cone defiend by these constraints. Tables 6 and 7 report both properties for the technology and the income-induced transition, respectively.

	H_0	MMM (least favorable) $\hat{T}_{\text{MMM}}$	$c_{0.05}^{\text{LF}}$	rej.	MMM (RSW) $c_{0.05}^{\text{RSW}}$	rej.	Fang–Seo projection $\hat{T}_{\text{FS}}$	$\hat{c}_{0.05}$ (p)	rej.
ages 0–5	convex complementarity	8.7	50.6	no	51.6	no	4.1	11.2 (0.57)	no
	concave substitutability	13.1	43.2	no	44.7	no	8.9	12.2 (0.27)	no
ages 6–13	convex complementarity	36.7	40.9	no	41.8	no	19.4	29.9 (0.10)	no
	concave substitutability	0.0	43.2	no	44.1	no	0.0	18.4 (1.00)	no

Table 6: Tests of convex-ordinal complementarity (Definition 3: $E[Q(\theta')]$ supermodular for every non-decreasing convex Q; H_0: the suffix sums $T_j(\theta, I) = \sum_{k \geq j} \Delta^2 \overline{F}(k \mid \theta, I)$ are all non-negative, Proposition 7) and concave-ordinal substitutability (H_0: the prefix sums $L_j(\theta, I) = \sum_{k \leq j} \Delta^2 \overline{F}(k \mid \theta, I)$ are all non-positive) for the technology $\overline{\Pi}(\theta' \mid \theta, I)$, by transition group. Same statistics, critical values, and tuning as Table 2, applied to the cumulative (suffix- or prefix-summed) cross-differences. Neither property is rejected by any test.

	H_0	MMM (least favorable) $\hat{T}_{\text{MMM}}$	$c_{0.05}^{\text{LF}}$	rej.	MMM (RSW) $c_{0.05}^{\text{RSW}}$	rej.	Fang–Seo projection $\hat{T}_{\text{FS}}$	$\hat{c}_{0.05}$ (p)	rej.
ages 0–5	convex complementarity	4.5	52.5	no	43.8	no	0.7	4.2 (0.82)	no
	concave substitutability	32.9	46.3	no	44.6	no	4.1	5.6 (0.07)	no
ages 6–13	convex complementarity	29.2	47.9	no	49.7	no	16.0	17.9 (0.07)	no
	concave substitutability	0.2	48.3	no	49.8	no	0.0	14.1 (1.00)	no

Table 7: Convex-ordinal complementarity and concave-ordinal substitutability of the income-induced transition $\Pi^{\text{ind}}(\theta' \mid \theta, y)$ in (10): same statistics, constraint sets, critical values, and tuning as Table 6, with income in place of investment. Neither property is rejected by any test.

Neither shape-refined property is rejected, for either object or stage, by any of the three tests—including the Fang–Seo projection test, which avoids least-favorable conservatism: the estimated technology is statistically consistent both with convex complementarity (every top-weighted valuation seeing complements) and with concave substitutability (every bottom-weighted valuation seeing substitutes). This is the formal counterpart of the anchor-dependence found throughout—which sign a planner sees depends on the curvature of the valuation, and the data refute neither curvature's verdict. The one place the shape tests come close to biting is school age, where the convex-complementarity statistic (37 for the technology, 29 induced) sits just below its least-favorable critical value and the corresponding Fang–Seo p-values fall to 0.10 and 0.07: even valuations that load on the top of the skill distribution begin to see substitutability there.

6.7 Counterfactuals

The estimated technology supports three policy experiments, in increasing order of realism. The first reallocates *investment* directly, holding its distribution fixed—the fixed-marginal planner problem (P1) of Section 3. The second reallocates *income* across children, letting investment respond through the estimated behavioural rule. A third, most realistic experiment—a budget-constrained *transfer* schedule on each household's own income—is developed in Appendix C. Each is evaluated for a set of *anchors*—adult outcomes whose expectation given terminal skill plays the role of the valuation Q from Section 3—namely $Q \in \{\text{HS}, \text{college}, \text{earnings}, \text{log earnings}, \text{employment}\}$. Each problem reduces to a (stacked) linear or low-dimensional program once the surplus is computed from Q and the estimated transitions, and is solved exactly.

A recurring feature of the results is that the optimal policy is nearly the same across anchors, with college completion the exception. Figure 4 shows why: it plots each anchor $Q(\theta_T) = E[\text{outcome} \mid \theta_T]$ as a function of terminal skill. High-school completion, employment, and (log) earnings are close to linear in θ_T, so their marginal value of a skill step is roughly constant and the planner weights skill gains similarly. College completion is sharply convex—almost all of its value sits in the top skill step—so its marginal-value profile is concentrated at the top, which tilts the college planner's allocation relative to the others.

The mechanism behind this agreement is precise. The planner's surplus for anchor Q is $E[Q(\theta_{t+1}) \mid \theta_t, I_t]$, and its local cross-difference decomposes, by summation by parts, as

$$\Delta^2 E[Q](\theta, I) = \sum_k \Delta Q(k)\, \Delta^2 \overline{F}_{Y|X}(k \mid \theta, I), \tag{11}$$

where $\Delta Q(k) = Q(k+1) - Q(k) \geq 0$ and $\Delta^2 \overline{F}_{Y|X}(k \mid \cdot)$ is the per-cutpoint survival difference-in-difference mapped in Figure 3. Each anchor's local complementarity is therefore the same set of survival difference-in-differences, reweighted by the anchor's marginal value of a skill step (Figure 4). Where the technology is uniformly super- or sub-modular across cutpoints, every monotone Q inherits the same sign—ordinal complementarity—and the allocation is anchor-invariant; where the cutpoint signs are mixed, the assignment depends on the relative magnitudes of $\Delta^2 E[Q]$ rather than its sign pattern alone. Across the estimated technology the five anchors share almost the same sign map, and college completion is the exception: because its value concentrates in the top skill step, its surplus places relatively more weight on the high-skill cells and is the one whose difference-in-differences are least proportional to the rest. That non-proportionality—not a sign reversal—is what separates college's optimal

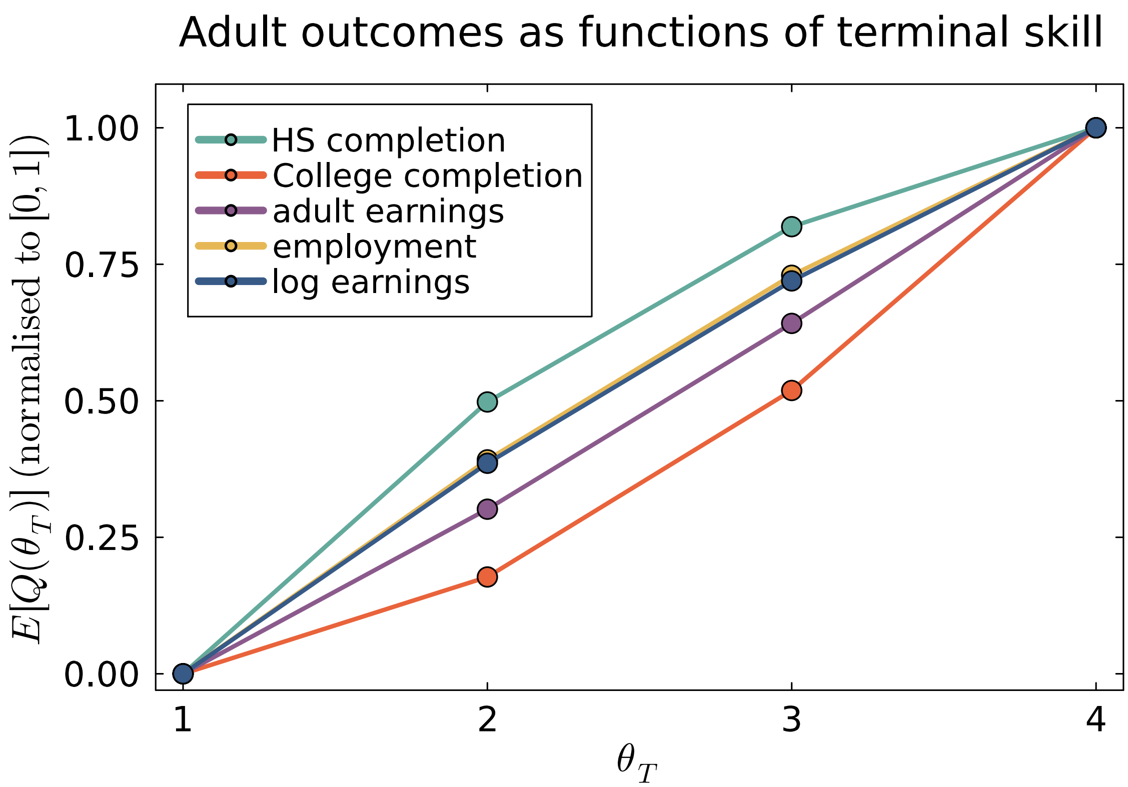

Figure 4: Adult outcomes as functions of terminal skill θ_T: each anchor $E[Q(\theta_T)]$, normalized to $[0, 1]$ to compare shapes. College completion is the one anchor whose value is concentrated in the top skill step; the rest are close to linear.

policy from the others.

6.7.1 Reallocating investment

The first experiment is the dynamic fixed-marginal problem of Section 3.2, solved over the full seven-period panel. Only the initial skill distribution F_{θ_1} is fixed at its empirical value; at each of the six investment periods $t = 1, \ldots, 6$ the planner re-couples investment I_t to current skill θ_t—holding the empirical investment marginal F_{I_t} fixed—and the resulting skill distribution is carried forward through the estimated technology:

$$\max_{\gamma_1,\ldots,\gamma_6} E[Q(\theta_T)] \quad \text{s.t.} \quad \gamma_t \in \Gamma(F_{\theta_t}, F_{I_t}), \qquad F_{\theta_{t+1}} = \textstyle\sum_{\theta,I} \Pi(\cdot \mid \theta, I)\, \gamma_t(\theta, I), \qquad F_{\theta_1} \text{ empirical.} \tag{12}$$

Investment is neither created nor destroyed—only re-assigned across children within each period—so the experiment isolates the technology's allocative content. Because the continuation value is convex and piecewise-linear in the skill distribution, (12) collapses to a single stacked linear program in $(\gamma_1, \ldots, \gamma_6)$ and is solved exactly (Section 3.2). Figure 5 contrasts, period by period, the data profile of $E[I \mid \theta]$ with the comonotone benchmark that ordinal complementarity would make optimal, and with the profiles the planners actually prescribe.

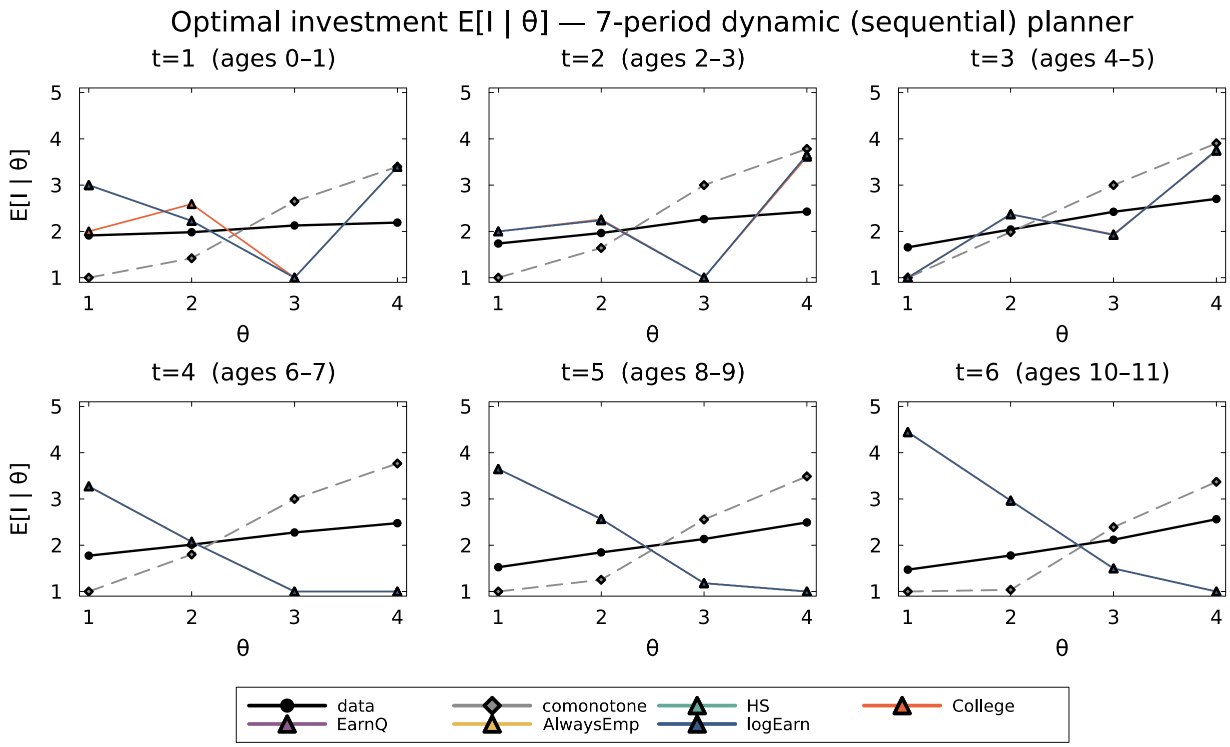

Figure 5: Optimal investment $E[I \mid \theta]$ under the full seven-period dynamic planner, by developmental period. The data profile is mildly increasing and the comonotone benchmark implied by ordinal complementarity is steeply increasing; from age six on the planners' allocations are sharply *anti*-comonotone (more investment to lower-skill children), while in early childhood—where the technology is mixed and on the boundary of ordinal complementarity—the optimal profile is non-monotone. In every period the five anchors nearly coincide.

The optimal allocations are decisively *anti*-comonotone at school age—higher-skill children receive *less* investment, the opposite of the comonotone benchmark—reflecting the dominance of locally substitutable regions there. In early childhood, where the technology is mixed and on the boundary of ordinal complementarity (Section 6.4), the optimal profile is non-monotone and only weakly anti-assortative. The terminal skill distribution shepherds the bottom of the distribution upward without pushing the top (Figure 6).

The near-coincidence of the five anchors in Figure 5 is itself a consequence of the dynamics. From ages six to thirteen the technology is substitution-dominant and the optimal

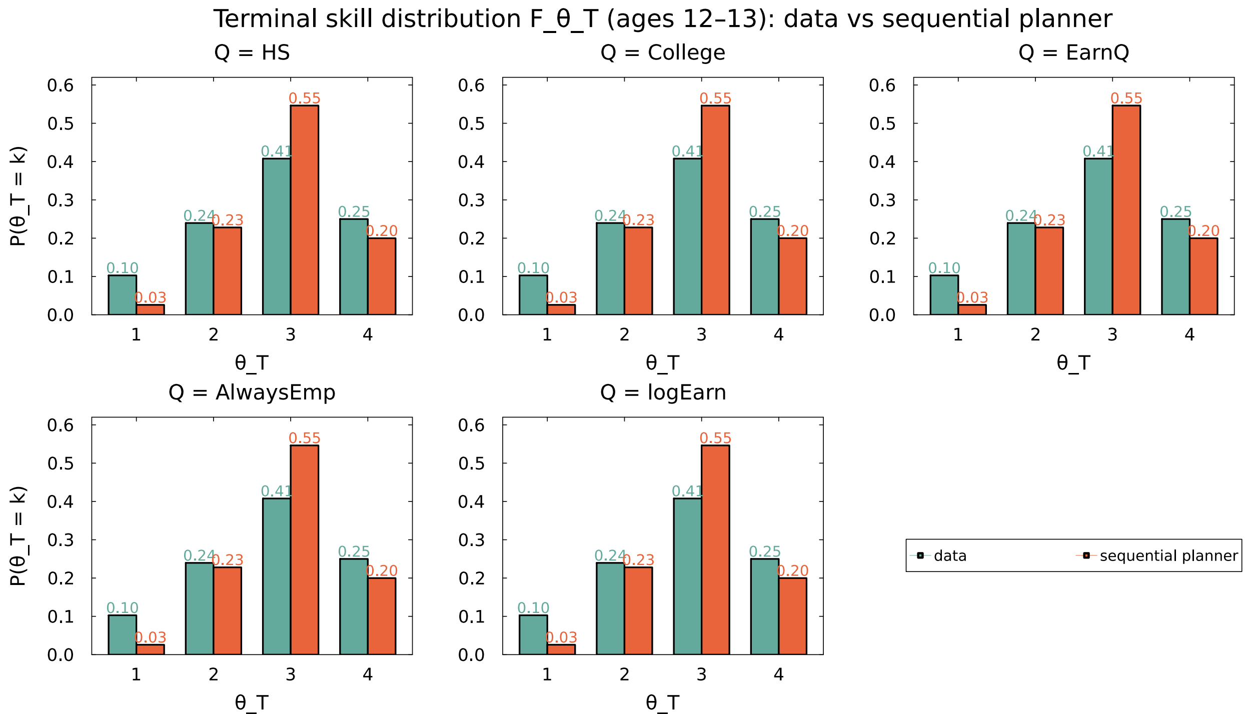

Figure 6: Terminal skill distribution under the seven-period dynamic planner, by anchor: the mass at $\theta_T = 1$ collapses while the middle rises, and the top is not pushed.

coupling is uniformly counter-monotone, so *every* monotone valuation—whatever its cardinal shape—prescribes the same anti-assortative allocation; there is nothing left for the anchor to disagree about. The agreement is therefore near-perfect at school age and frays only in early childhood, where the boundary technology makes the allocation sensitive to the anchor's curvature. Crucially, this reconciliation is a property of the *dynamic* solution. A myopic planner that re-optimizes period by period on the empirical marginals—valuing next-period skill directly by Q rather than by the continuation value—produces early-childhood allocations that both differ sharply from the dynamic optimum and no longer coincide across anchors, precisely because early childhood is where the horizon matters; from school age onward the myopic and dynamic solutions are almost identical (Appendix B, Figure 9). Looking through to the terminal outcome, and internalizing that school age is substitution-dominant, is what pulls the early-childhood allocations into line.

6.7.2 Reallocating income

A more policy-relevant experiment fixes the *behavioural* investment rule and reallocates income across children by initial skill θ_0, letting investment respond through the estimated

CCP. The planner chooses a per-skill income level $y(\theta_0)$ subject to a per-capita budget—the average permanent income in the population—weighting each skill level by its population share $p(\theta_0)$:

$$\max_{y(\cdot)} \; E_{\theta_0}\Big[E[Q(\theta_T) \mid \theta_0]\Big] \quad \text{s.t.} \quad \sum_{\theta_0} p(\theta_0)\, y(\theta_0) = E_{\theta^P,\omega,\varepsilon_Y}[Y(\theta^P,\omega,\varepsilon_Y)],$$

where $p(\theta_0)$ is the marginal distribution of initial skill. The transmission runs from $\Delta \log y$ to the CCP index to $\mathbb{P}(I_t \mid \cdot)$ to the technology; the log-income semi-elasticity of the CCP index is $\hat{\beta}_5/\hat{\delta} = 3.10$ (child-bootstrap SE 0.28, against a delta-method SE of 0.06). The optimal income profile is *non-monotone* in θ_0 (Figure 7): the lowest-skill children receive the most, but the profile then dips and rises again, with a secondary peak at $\theta_0 = 3$ rather than declining monotonically. The shape is most pronounced for the college anchor—the mixed, non-substitutable early-childhood region (Section 6.4) means the optimal income allocation turns on the cardinal shape of Q, not a single sign. The terminal skill distribution shifts accordingly (Figure 8).

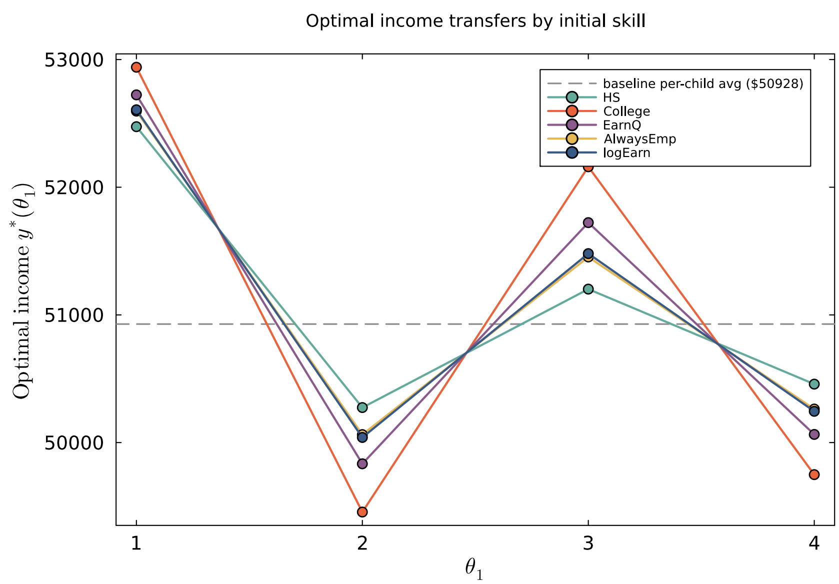

Figure 7: Optimal income given initial skill, by anchor. The profile is non-monotone in θ_0: highest at the lowest skill, with a secondary peak at $\theta_0 = 3$. College is the most pronounced.

A third and most realistic experiment leaves each household's own income in place and adds a budget-constrained transfer schedule before investment is chosen. Its optimal schedule is, like the income reallocation, *non-monotone*: transfers peak at the $15,000 income knot rather than at the very poorest households, and fall to zero further up. As with the income

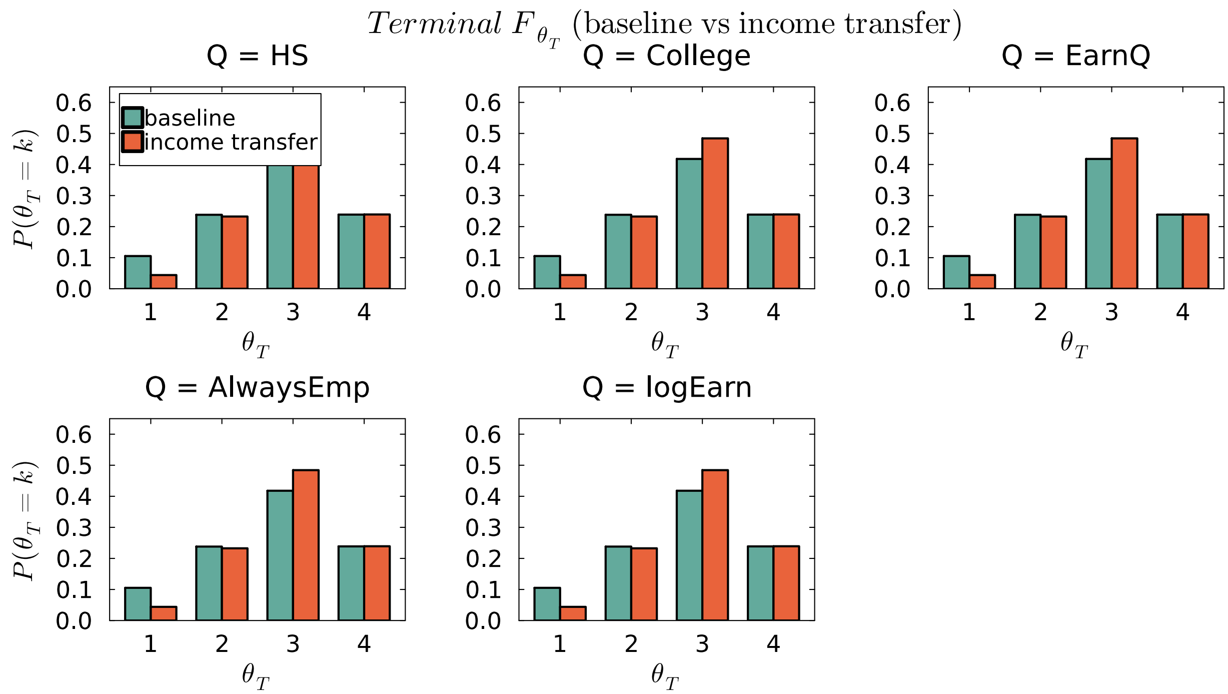

Figure 8: Terminal skill distribution under the optimal income reallocation, by anchor (baseline vs. income-transfer optimum).

Outcome	W_{base}	W_{opt}	Δ (%)
HS completion	0.921	0.930	+1.0
College completion	0.258	0.273	+5.8
Adult earnings quintile	25,455	26,159	+2.8
Employment 25–30	0.768	0.780	+1.6

Table 8: Welfare gains from the optimal reallocation of log income, by anchor, at the regularized fit. Unlike the investment experiment, this is a clean optimization against a like-for-like baseline (W_{base} fixes the behavioural CCP at the observed income marginal), so the gains are positive.

reallocation, this non-monotonicity originates in the mixed early-childhood technology—at school age the substitution-dominant technology alone would call for a monotone progressive transfer, and it is the boundary early-childhood region that bends the profile. Appendix C states the problem precisely and reports the optimal schedules and welfare gains by anchor.

Across the experiments the optimal policy is broadly redistributive toward lower-skill, lower-income children—though non-monotone in the boundary early-childhood region—and the choice of anchor affects its magnitude far more than its shape. Both features trace to the estimated technology: the prevalence of locally substitutable regions makes redistribution efficient, and (away from the boundary of ordinal complementarity) the sign of the optimal assignment is governed by the survival function rather than the particular cardinal valuation, leaving college—whose value loads on the top skill step—as the one consistent outlier.

7 Conclusion

Ordinal complementarity provides a scale-free definition of complementarity between skills and investment, robust to the choice of anchor and equivalent to supermodularity of $E[Q(Y) \mid X]$ for every monotone Q. It is necessary and sufficient for a broad class of static and dynamic planner problems to have monotone, assortative solutions, and it is testable through moment inequalities. Applied to the technology of cognitive skill formation in the NLSY Children sample, under conservative child-bootstrap inference the property is rejected at school age but not in early childhood, and the estimated technology is neither globally complementary nor globally substitutable: locally super- and sub-modular regions coexist, with the two developmental stages bracketing them—early childhood consistent with complementarity and robustly non-substitutable, school age substitution-dominant. The planner's optimal response is correspondingly non-monotone—progressive transfers and counter-monotone investment—and is largely invariant to the adult-outcome anchor used to evaluate it.

A Proofs

Proof of Proposition 1. Because G, g_1, g_2 are one-to-one on their ranges, the "if" direction is the "only if" direction applied to the inverses, so it suffices to prove "only if." Fix x, x', y with $x_j = x'_j$ for $j > 2$ and set $\tilde{y} = G(y)$, $\tilde{x} = (g_1(x_1), g_2(x_2), x_3, \dots)$, $\tilde{x}' = (g_1(x'_1), g_2(x'_2), x_3, \dots)$. Monotonicity of g_1, g_2 makes the min/max operations commute with the relabelling, so $F_{\tilde{Y}|\tilde{X}}(\tilde{y} \mid \tilde{x}) = F_{Y|X}(y \mid x)$ and likewise for x', $x \wedge x'$, $x \vee x'$. The inequality (4) transfers verbatim. $\square$

Proof of Proposition 2. Fix x, x' with $x_j = x'_j$ for $j > 2$ and define measures on $\mathcal{Y}$ by $\mu((-\infty, y]) = \frac{1}{2}[F_{Y|X}(y \mid x \wedge x') + F_{Y|X}(y \mid x \vee x')]$ and $\nu((-\infty, y]) = \frac{1}{2}[F_{Y|X}(y \mid x) + F_{Y|X}(y \mid x')]$. Definition 1 is exactly μ FOSD ν, and $\int Q\, d\mu \geq \int Q\, d\nu$ for every non-decreasing integrable Q by the standard stochastic-dominance characterization (Shaked and Shanthikumar, 2007, Eq. (1.A.7)). Rearranging gives (5); conversely, taking $Q = \mathbf{1}\{\cdot > \bar{y}\}$ recovers (4). $\square$

Additive models. The next three results support Section 2.3. Throughout, $Y = f(X) + \varepsilon$ with $\varepsilon \perp X$, f supermodular and strictly increasing with strictly positive first partial derivatives, H the CDF of ε and h its density, so that $F_{Y|X}(y \mid x) = H(y - f(x))$ and $\overline{F}_{Y|X}(y \mid x) = 1 - H(y - f(x))$. Ordinal complementarity in (x_1, x_2) is supermodularity of $\overline{F}_{Y|X}$ in those coordinates.

Lemma 1 (Additive criteria)**.** *(a) If H and f are continuously differentiable, ordinal complementarity holds if and only if $h(y - f(x))\,\partial_{x_1} f(x)$ is non-decreasing in x_2 for all y, x (equivalently, $h(y - f(x))\,\partial_{x_2} f(x)$ is non-decreasing in x_1).*

(b) If H and f are twice continuously differentiable, ordinal complementarity holds if and only if, for all y, x,

$$\partial_{x_1} f(x)\,\partial_{x_2} f(x) h'(y - f(x)) \;\le\; h(y - f(x))\,\partial^2_{x_1 x_2} f(x).$$

Proof. *(a)* For a differentiable function on a product of intervals, supermodularity is equivalent to $\partial_{x_1}\overline{F}_{Y|X}$ being non-decreasing in x_2. Since $\partial_{x_1}\overline{F}_{Y|X}(y \mid x) = h(y - f(x))\,\partial_{x_1} f(x)$, this is exactly the stated condition; the parenthetical is the symmetric cross-derivative. *(b)* With H, f twice differentiable, supermodularity is $\partial^2_{x_1 x_2}\overline{F}_{Y|X} \ge 0$. Differentiating $\partial_{x_1}\overline{F}_{Y|X} = h(y - f)\,\partial_{x_1} f$ in x_2,

$$\partial^2_{x_1 x_2}\overline{F}_{Y|X}(y \mid x) = h(y - f)\,\partial^2_{x_1 x_2} f - h'(y - f)\,\partial_{x_1} f\,\partial_{x_2} f.$$

Hence $\partial^2_{x_1 x_2}\overline{F}_{Y|X}(y \mid x) \ge 0$ if and only if

$$\partial_{x_1} f(x)\,\partial_{x_2} f(x) h'(y - f(x)) \;\le\; h(y - f(x))\,\partial^2_{x_1 x_2} f(x)$$

□

Proposition 8 (Failure under thin tails)**.** *Suppose H and f are twice continuously differentiable and the score $h'(z)/h(z)$ is unbounded above. Then ordinal complementarity fails. In particular it fails for every Gaussian shock, establishing case (i) of Section 2.3.*

Proof. Fix any x. The right-hand side of Lemma 1(b), $\partial^2_{x_1 x_2} f(x)/(\partial_{x_1} f(x)\,\partial_{x_2} f(x))$, is a finite constant, while as y ranges over $\mathbb{R}$ the argument $z = y - f(x)$ ranges over $\mathbb{R}$ and the left-hand side $h'(z)/h(z)$ takes arbitrarily large values. The criterion of Lemma 1(b) therefore fails at some y, so $\overline{F}_{Y|X}$ is not supermodular. For $\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$ the score is $h'(z)/h(z) = -(z - \mu)/\sigma^2$, which is unbounded above as $z \to -\infty$; the same holds whenever the left tail of h decays faster than every exponential, i.e. $h'(z)/h(z) \to \infty$ as $z \to -\infty$. Case (ii) is separate: a shock with bounded support makes the conditional support of Y depend on (x_1, x_2), which already violates (4) for y near a support boundary. $\square$

[A heavy-tailed shock for which ordinal complementarity holds] Let $f(x_1, x_2) = u(x_1)v(x_2)$ with strictly increasing $u, v : \mathbb{R} \to (0, 1)$ and strictly positive first derivatives (for example, u and v may be cdfs with full support in $\mathbb{R}$ and no mass points) , and let $Y = f(X) + \log \eta$ with $\eta \sim \text{Exp}(\lambda)$ independent of X, so that $-\varepsilon$ is a Gumbel-type (reflected extreme-value) shock. Then $H(z) = 1 - \exp(-\lambda e^z)$, $h(z) = \lambda e^z \exp(-\lambda e^z)$, and $h'(z)/h(z) = 1 - \lambda e^z$. Writing $w = y - f(x)$, the cross-partial of the conditional CDF is

$$\partial^2_{x_1 x_2} F_{Y|X}(y \mid x) = u'(x_1)\, v'(x_2)\, h(w) \Big[(1 - \lambda e^w)\, u(x_1)v(x_2) - 1\Big] < 0,$$

since $uv < 1$ and $1 - \lambda e^w < 1$ force the bracket strictly below zero. Hence $F_{Y|X}$ is strictly submodular, $\overline{F}_{Y|X}$ strictly supermodular, and ordinal complementarity holds for every such f—no matter how steep the supermodular interaction $u'v'$, the heavy left tail of the shock dominates. The Laplace and Logistic cases (iii) are analogous, with criterion (b) satisfied once the scale σ exceeds the range A of f.

Proposition 9 (Multivariate equivalences)**.** *With $Y = (Y_1, \ldots, Y_r)$: (1) X_1, X_2 are weakly ordinal complements in Y iff $g_1(X_1), g_2(X_2)$ are weakly ordinal complements in $(G_1(Y_1), \ldots, G_r(Y_r))$ for all increasing G_j, g_1, g_2, iff $E[\prod_j Q_j(Y_j) \mid \cdot]$ is supermodular for all non-negative increasing Q_j. (2) The strong-sense analogues hold with $\prod_j Q_j(Y_j)$ replaced by any increasing integrable $Q : \mathbb{R}^r \to \mathbb{R}$. Moreover, if X_1, X_2 are ordinal complements in each Y_j and the conditional copula of Y given X does not depend on (x_1, x_2), then they are strong-sense ordinal complements in Y.*

Proof. The measures μ, ν of the previous proof, now on $\mathbb{R}^r$, place the claim within multivariate stochastic-order theory: the weak sense is upper-orthant dominance and the strong sense is the usual stochastic order (Shaked and Shanthikumar, 2007, Ch. 6); the cited theorems (6.G.1, 6.G.3 for the weak case; 6.B.4, 6.B.16 for the strong case; 6.B.14 for the copula sufficient condition) give the stated equivalences. $\square$

Proof of Corollary 1. For any coupling γ and cutpoint $\bar{y}$, the implied outcome marginal satisfies $\mathbb{P}_\gamma(Y > \bar{y}) = \iint \overline{F}_{Y|X}(\bar{y} \mid x)\, d\gamma$, the welfare of γ under $Q_{\bar{y}} = \mathbf{1}\{\cdot > \bar{y}\}$. (*Only if.*) Under ordinal complementarity the surplus $\overline{F}_{Y|X}(\bar{y} \mid \cdot)$ is supermodular, so the comonotone coupling maximizes $\mathbb{P}_\gamma(Y > \bar{y})$ over $\Gamma(F_{X_1}, F_{X_2})$ (Proposition 3); varying $\bar{y}$ gives FOSD. (*If.*)

If the comonotone outcome marginal dominates at every pair of marginals, then for each $\bar{y}$ the comonotone coupling maximizes $\mathbb{P}_\gamma(Y > \bar{y})$ at every pair of marginals, so Proposition 3 makes $\overline{F}_{Y|X}(\bar{y} \mid \cdot)$ supermodular; as $\bar{y}$ ranges over $\mathcal{Y}$ this is Definition 1. $\square$

Proof of Proposition 6. Sufficiency follows by noting that if $t^+(F) \geq 0$ then the solution ΔQ that delivers such a a value must also, by definition, satsify the necessary inequalities for supermodularity. A Q can be construct simply by setting $Q(k) = \sum_{j<k} \Delta Q(j)$. For necessity, suppose that such a Q exists and define $\Delta\tilde{Q}(k) = \Delta Q(k)/\sum_j \Delta Q(j)$. Since $\Delta\tilde{Q} \in \mathcal{Q}$, by definition $t^+(F) \geq \min_{x_1,x_2} \sum_k \Delta\tilde{Q}(k), \Delta^2\overline{F}_{Y|X}(k \mid x_1, x_2) \geq 0$. $\square$

Proof of Proposition 7. Apply summation by parts to (8): with $T_j = \sum_{k\geq j} \Delta^2\overline{F}_{Y|X}(k \mid \cdot)$,

$$\Delta^2 E[Q](x_1, x_2) = \Delta Q(1)\, T_1(x_1, x_2) + \sum_{j\geq 2} \Big(\Delta Q(j) - \Delta Q(j-1)\Big)\, T_j(x_1, x_2).$$

Sufficiency. A non-decreasing convex Q has $\Delta Q(1) \geq 0$ and $\Delta Q(j) - \Delta Q(j-1) \geq 0$; if every $T_j \geq 0$ on every cell, the right side is non-negative, so the surplus is supermodular for every such Q—convex-ordinal complementarity. *Necessity.* Fix a cutpoint j_0 and a cell. The increments $\Delta Q(k) = \mathbf{1}\{k \geq j_0\}$ define a non-decreasing convex Q (its increments are 0 then 1, non-decreasing), and for it $\Delta^2 E[Q](x_1, x_2) = T_{j_0}(x_1, x_2)$. Hence convex-ordinal complementarity forces $T_{j_0} \geq 0$ at every cell, for every j_0. The concave/prefix-sum statement is the mirror image, with $L_j = \sum_{k\leq j} \Delta^2\overline{F}_{Y|X}(k \mid \cdot)$ and $\Delta Q(k) = \mathbf{1}\{k \leq j_0\}$. $\square$

The budget problem. Proposition 3 is stated without proof; the budget-problem counterpart is proved here.

Proof of Proposition 4. (*Only if.*) The cost of a coupling depends on it only through its X_2-marginal. Given any solution γ of (P2) with induced marginal $F^\gamma_{X_2}$, the comonotone coupling of $(F_{X_1}, F^\gamma_{X_2})$ has the same cost, hence is feasible, and attains weakly higher welfare by Proposition 3; it is therefore a comonotone solution.

(*If.*) Suppose q is not supermodular: $q(x \wedge x') + q(x \vee x') < q(x) + q(x')$ for some pair x, x'. The strict inequality forces $x_1 \neq x'_1$ and $x_2 \neq x'_2$—otherwise $\{x \wedge x', x \vee x'\} = \{x, x'\}$ and the two sides coincide—so relabelling the pair supplies $x_1 < x'_1$ and $x'_2 < x_2$ with

$$\Delta_L := q(x_1, x_2) - q(x_1, x'_2) \;>\; q(x'_1, x_2) - q(x'_1, x'_2) =: \Delta_H,$$

the two types' returns to raising investment from x'_2 to x_2; Assumption 1 makes q non-decreasing in x_2, so $\Delta_L > \Delta_H \geq 0$. Let F_{X_1} place mass $\frac{1}{2}$ on each of x_1, x'_1, set $B = \frac{1}{2}$, let

$q_{\text{lower}} = \min\{q(x_1, x_2'), q(x_1', x_2')\}$, and choose the cost

$$b(v) = \begin{cases} 0 & v \le x_2', \\ 1 & x_2' < v \le x_2, \\ \dfrac{\max\{q(x_1,v),\ q(x_1',v)\} - q_{\text{lower}}}{\Delta_L} + 1 & v > x_2, \end{cases}$$

which is non-decreasing by Assumption 1 (monotonicity of $q(x_1,\cdot)$ and $q(x_1',\cdot)$) and satisfies, for $v > x_2$ and $x \in \{x_1, x_1'\}$,

$$\Delta_L\, b(v) \ > \ q(x,v) - q_{\text{lower}}. \tag{$\star$}$$

By monotonicity of q in x_2, any feasible coupling γ is weakly dominated, at identical cost, by the coupling $\tilde\gamma$ that moves its mass at levels $v \le x_2'$ up to x_2' (zero cost either way) and its mass at $v \in (x_2', x_2)$ up to x_2 (unit cost either way), leaving mass above x_2 unchanged. Write $a = \tilde\gamma(x_1, x_2)$, $c = \tilde\gamma(x_1', x_2)$, and $W_0 = \frac{1}{2}q(x_1,x_2') + \frac{1}{2}q(x_1',x_2')$; the budget reads $a + c + \int_{(x_2,\infty)} b\, d\tilde\gamma(x_1,\cdot) + \int_{(x_2,\infty)} b\, d\tilde\gamma(x_1',\cdot) \le \frac{1}{2}$, i.e.

$$a + c \ \le \ \tfrac{1}{2} - \int_{(x_2,\infty)} b\, d\tilde\gamma(x_1,\cdot) - \int_{(x_2,\infty)} b\, d\tilde\gamma(x_1',\cdot).$$

Defining the non-negative quantities (by ($\star$))

$$I_1 := \int_{(x_2,\infty)} \Big[\Delta_L b(v) - (q(x_1,v) - q_{\text{lower}})\Big] d\tilde\gamma(x_1,v), \qquad I_2 := \int_{(x_2,\infty)} \Big[\Delta_L b(v) - (q(x_1',v) - q_{\text{lower}})\Big] d\tilde\gamma(x_1',v),$$

strictly positive unless $\tilde\gamma(x_1,\cdot), \tilde\gamma(x_1',\cdot)$ respectively vanish on (x_2,∞), we obtain

$$\mathcal{W}(\gamma) \ \le \ \mathcal{W}(\tilde\gamma) \ \le \ W_0 + a\Delta_L + c\Delta_H + \int_{(x_2,\infty)} \Big[q(x_1,\cdot) - q_{\text{lower}}\Big] d\tilde\gamma(x_1,\cdot) + \int_{(x_2,\infty)} \Big[q(x_1',\cdot) - q_{\text{lower}}\Big] d\tilde\gamma(x_1',\cdot)$$

$$\le \ (a+c)\Delta_L + \int_{(x_2,\infty)} \Big[q(x_1,\cdot) - q_{\text{lower}}\Big] d\tilde\gamma(x_1,\cdot) + \int_{(x_2,\infty)} \Big[q(x_1',\cdot) - q_{\text{lower}}\Big] d\tilde\gamma(x_1',\cdot) + W_0$$

$$\le \ W_0 + \tfrac{1}{2}\Delta_L - I_1 - I_2 \ \le \ W_0 + \tfrac{1}{2}\Delta_L.$$

The bound is attained by the feasible coupling $\gamma^\star = \frac{1}{2}\delta_{(x_1,x_2)} + \frac{1}{2}\delta_{(x_1',x_2')}$, so (P2) admits a solution. At any solution every inequality holds with equality, forcing $c = 0$ (third inequality, using $\Delta_H < \Delta_L$), $I_1 = I_2 = 0$ (final equality), hence $\tilde\gamma(x_1,\cdot), \tilde\gamma(x_1',\cdot)$ vanish on (x_2,∞), and $a + c = \frac{1}{2}$ (fourth and fifth inequality, budget tight since $\Delta_L > 0$), hence $a = \frac{1}{2}$: undoing the dominating moves, every solution places all type-x_1 mass at levels in $(x_2', x_2]$—exhausting the budget on the low-skill type—and all type-x_1' mass at levels $\le x_2'$. Every solution therefore assigns the low-skill type strictly higher investment than the high-skill type, and none is comonotone. $\square$

Dynamic results.

Proof of Proposition 5. Let $\{F^*_{I_t|\theta_t}\}_{t=0}^T$ solve (6), and write γ^*_t for the joint law of (θ_t, I_t) it induces, $F^*_{\theta_t}$ and $F^*_{I_t}$ for the induced marginals, so that $\gamma^*_t \in \Gamma(F^*_{\theta_t}, F^*_{I_t})$ and $F^*_{\theta_{t+1}} = \int F_{t+1}(\cdot \mid \theta, I)\, d\gamma^*_t$.

Define a candidate policy recursively. Set $F^{**}_{\theta_0} = F_{\theta_0}$, and given $F^{**}_{\theta_t}$ let γ^{**}_t be the comonotone coupling of $\left(F^{**}_{\theta_t}, F^*_{I_t}\right)$, that is, the law of

$$\left(\left(F^{**}_{\theta_t}\right)^{-1}(U),\ \left(F^*_{I_t}\right)^{-1}(U)\right), \qquad U \sim \mathrm{Unif}(0,1),$$

and set $F^{**}_{\theta_{t+1}} = \int F_{t+1}(\cdot \mid \theta, I)\, d\gamma^{**}_t$. Equivalently, $I^{**}_t = \left(F^*_{I_t}\right)^{-1} \circ F^{**}_{\theta_t}(\theta_t)$. The three steps below show that this policy is feasible, that it induces a stochastically larger skill path, and hence that it is optimal.

Step 1: feasibility. By construction γ^{**}_t has second marginal $F^*_{I_t}$, so the induced investment marginal is unchanged, and the budget constraint—which depends on the policy only through that marginal—reads

$$\int b_t\, dF^{**}_{I_t} = \int b_t\, dF^*_{I_t} \ \leq\ B_t.$$

The candidate is therefore feasible, and it remains so verbatim if the budget is imposed with equality.

Step 2: the candidate skill path dominates. We show by induction that $F^{**}_{\theta_t}$ FOSD $F^*_{\theta_t}$ for every t. At $t = 0$ the two coincide. Assume the claim at t, and let $\tilde{\gamma}_t$ denote the comonotone coupling of $\left(F^*_{\theta_t}, F^*_{I_t}\right)$—the same investment marginal as γ^{**}_t, but the *old* skill marginal. We compare $F^{**}_{\theta_{t+1}}$ to $F^*_{\theta_{t+1}}$ through this intermediate object.

First, γ^*_t and $\tilde{\gamma}_t$ have identical marginals, so Corollary 1 applies: under ordinal complementarity the comonotone coupling of a given pair of marginals induces an outcome distribution that first-order stochastically dominates the one induced by any other coupling of those marginals. Hence

$$\int F_{t+1}(\cdot \mid \theta, I)\, d\tilde{\gamma}_t \ \ \mathrm{FOSD} \ \ \int F_{t+1}(\cdot \mid \theta, I)\, d\gamma^*_t = F^*_{\theta_{t+1}}.$$

Second, γ_t^{**} and $\tilde{\gamma}_t$ are both comonotone and share the second marginal $F^*_{I_t}$, so by the quantile representation of the comonotone coupling, for every z,

$$\overline{F}^{**}_{\theta_{t+1}}(z) = \int_0^1 \overline{F}_{t+1}\Big(z \;\Big|\; \big(F^{**}_{\theta_t}\big)^{-1}(u),\ \big(F^*_{I_t}\big)^{-1}(u)\Big)du,$$

and likewise for $\tilde{\gamma}_t$ with $\big(F^*_{\theta_t}\big)^{-1}$ in place of $\big(F^{**}_{\theta_t}\big)^{-1}$. The induction hypothesis is equivalent to $\big(F^{**}_{\theta_t}\big)^{-1}(u) \geq \big(F^*_{\theta_t}\big)^{-1}(u)$ for all $u \in (0,1)$, and FOSD-monotonicity makes $\overline{F}_{t+1}(z \mid \cdot, I)$ non-decreasing in its first argument, so the integrand dominates pointwise in u. Hence $F^{**}_{\theta_{t+1}}$ FOSD $\int F_{t+1}(\cdot \mid \theta, I)\, d\tilde{\gamma}_t$.

Chaining the two displays and using transitivity of first-order stochastic dominance gives $F^{**}_{\theta_{t+1}}$ FOSD $F^*_{\theta_{t+1}}$, completing the induction.

Step 3: optimality. Each Q_t is non-decreasing and $F^{**}_{\theta_t}$ FOSD $F^*_{\theta_t}$, so $E_{F^{**}_{\theta_t}}[Q_t(\theta_t)] \geq E_{F^*_{\theta_t}}[Q_t(\theta_t)]$ for every t, and summing over t shows that the candidate attains a weakly higher objective than the solution $\{F^*_{I_t|\theta_t}\}$. Being also feasible by Step 1, it is itself a solution. Its period-t coupling is comonotone by construction, and since $F^{**}_{I_t} = F^*_{I_t}$ the comonotone rule may be written in terms of the marginals it induces, $I_t^{**} = \big(F^{**}_{I_t}\big)^{-1} \circ F^{**}_{\theta_t}(\theta_t)$. $\square$

It may be of interest to also consider the following problem:

$$\begin{aligned}
\sup_{\{B_t, A_{t+1}, F_{I_t|\theta_t}\}_{t=0}^T} \sum_{t=0}^{T+1} G_t(B_t) + \int_{\theta_t} Q_t(c_t) dF_{\theta_t}(c_t) \\
B_t + A_{t+1} = A_t(1+r_t) \\
A_{t+1} \geq \underline{A}_t \\
\int_{I_t}\int_{\theta_t} b_t(i_t) dF_{I_t|\theta_t}(i_t|c_t) dF_{\theta_t}(c_t) \leq B_t \text{ for all } t \\
F_{\theta_t}(c_t) = \int_{\theta_{t-1}}\int_{I_{t-1}} F_{\theta_t|\theta_{t-1},I_{t-1}}(c_t|c_{t-1}, i_{t-1}) dF_{I_{t-1}|\theta_{t-1}}(i_{t-1}) dF_{\theta_{t-1}}(c_{t-1}) \text{ for all } t \\
A_0 \text{ given}
\end{aligned}$$

If the previous problem has a solution, then it has a co-monotone solution.

Corollary 3. *Suppose that the previous problem has a solution $\{B_t^*, A_{t+1}^*, F^*_{I_t|\theta_t}\}_{t=0}^T$. Then, it has a solution $\{B_t^*, A_{t+1}^*, F^{**}_{I_t|\theta_t}\}_{t=0}^T$ in which the copula associated to $F^{**}_{I_t|\theta_t}$ is comonotone.*

Proof. Fix $\{B_t^*, A_t^*\}_{t=0}^T$ and consider the problem of choosing $\{F_{I_t|\theta_t}\}$ to maximize planner's objective. Since this problem has at least a solution (which is $\{F^*_{I_t|\theta_t}\}$), it has a comonotone solution by Proposition 5. Let this solution be $F^{**}_{I_t|\theta_t}$, and note that because it yields at least the payoff associated to $\{F^{*}_{I_t|\theta_t}\}_{t=0}^{T}$ given $\{B^*_t, A^*_{t+1}\}_{t=0}^{T}$ then $\{B^*_t, A^*_{t+1}, F^{**}_{I_t|\theta_t}\}_{t=0}^{T}$ gives at least the same payoff as $\{B^*_t, A^*_{t+1}, F^{*}_{I_t|\theta_t}\}_{t=0}^{T}$, so it must be a solution. This completes the proof. □

References

Francesco Agostinelli and Matthew Wiswall. Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. NBER Working Paper 22441, 2016.

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. *Journal of Political Economy*, 133(3):846–887, 2025.

D. W. K. Andrews and P. Guggenberger. Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory*, 25(3): 669–709, 2009.

Donald W. K. Andrews and Gustavo Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.

Susan Athey. Characterizing properties of stochastic objective functions. 1998.

Susan Athey. Monotone comparative statics under uncertainty. *Quarterly Journal of Economics*, 117(1):187–223, 2002.

Uta Bolt, Eric French, Jamie Hentall-MacCuish, and Cormac O'Dea. The intergenerational elasticity of earnings: exploring the mechanisms. *Journal of Political Economy*, Forthcoming, 2026.

S. Bonhomme and J.-M. Robin. Consistent noisy independent component analysis. *Journal of Econometrics*, 149(1):12–25, 2009.

Guillaume Carlier. On a class of multidimensional optimal transportation problems. *Journal of convex analysis*, 10(2):517–530, 2003.

Elizabeth Caucutt, Lance Lochner, Joseph Mullins, and Youngmin Park. Child skill production: Accounting for parental and market-based time and goods investments. *Journal of Political Economy*, 134(1):150–209, 2026.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104 (9):2633–79, September 2014. doi: 10.1257/aer.104.9.2633. URL https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2633.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Flavio Cunha and James J. Heckman. The technology of skill formation. *American Economic Review: Papers and Proceedings*, 97(2):31–47, 2007.

Zheng Fang and Juwon Seo. A projection framework for testing shape restrictions that form convex cones. *Econometrica*, 89(5):2439–2458, 2021.

Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 2016.

Sebastián Gallegos and Jorge Luis García. Childcare and parenting in the production of early life skills. *Economics of Education Review*, 101:102557, 2024. ISSN 0272-7757. doi: https://doi.org/10.1016/j.econedurev.2024.102557. URL https://www.sciencedirect.com/science/article/pii/S0272775724000517.

Martin Garcia-Vazquez. Identification and estimation of non-stationary hidden markov models. Working paper, 2021.

James J Heckman, Haihan Tian, Zijian Zhang, and Jin Zhou. Dynamic complementarity. Working Paper 34833, National Bureau of Economic Research, February 2026. URL http://www.nber.org/papers/w34833.

Johannes Hollenbach, Hendrik Schmitz, and Matthias Westphal. Gene-environment interactions with essential heterogeneity. *The Economic Journal*, page ueag010, 01 2026. ISSN 0013-0133. doi: 10.1093/ej/ueag010. URL https://doi.org/10.1093/ej/ueag010.

Y. Hu and S. M. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008.

Rucker C. Johnson and C. Kirabo Jackson. Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. *American Economic Journal: Economic Policy*, 11(4):310–49, November 2019. doi: 10.1257/pol.20180510. URL https://www.aeaweb.org/articles?id=10.1257/pol.20180510.

Josh Kinsler. Teacher complementarities in test score production: Evidence from primary school. *Journal of Labor Economics*, 34(1):29–61, 2016. doi: 10.1086/682331. URL https://www.journals.uchicago.edu/doi/abs/10.1086/682331.

Costas Meghir, Orazio Attanasio, Pamela Jervis, Monimalika Day, Prerna Makkar, Jere Behrman, Prachi Gupta, Rashim Pal, Angus Phimister, Nisha Vernekar, et al. Early stimulation and enhanced preschool: a randomized trial. *Pediatrics*, 151(Supplement 2): e2023060221H, 2023.

Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(1): 157–180, 1994.

Gloria Moroni, Cheti Nicoletti, and Emma Tominey. Stressful home environment and the child's socio-emotional development. *Journal of Human Capital*, 19(3):474–514, 2025. doi: 10.1086/736014. URL https://www.journals.uchicago.edu/doi/abs/10.1086/736014.

Dilnoza Muslimova, Hans van Kippersluis, Cornelius A. Rietveld, Stephanie von Hinke, and S. Fleur W. Meddens. Gene-environment complementarity in educational attainment. *Journal of Labor Economics*, 44(3):759–788, 2026. doi: 10.1086/734087. URL https://www.journals.uchicago.edu/doi/abs/10.1086/734087.

Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002, 2014.

Moshe Shaked and J. George Shanthikumar. *Stochastic Orders*. Springer, 2007 edition, 2007.

Donald M Topkis. Minimizing a submodular function on a lattice. *Operations research*, 26 (2):305–321, 1978.

B The myopic static planner

Section 6.7.1 noted that the anchor-invariance of the optimal investment allocation is a property of the *dynamic* solution. Figure 9 shows the allocation chosen instead by a myopic planner that, at each period t, solves a one-shot optimal-transport problem on the *empirical* marginals (F_{θ_t}, F_{I_t}) with the one-step reward $c_t(\theta, I) = E[Q(\theta_{t+1}) \mid \theta, I]$—valuing next-period skill directly by the adult outcome Q, with no continuation value and no forward chaining. From age six on the myopic and dynamic allocations are almost identical: the school-age technology is substitution-dominant, so both are uniformly counter-monotone and both are anchor-invariant. In early childhood they diverge sharply, and the myopic anchors no longer coincide with one another—the mixed, boundary technology there makes the one-step allocation sensitive to the cardinal shape of Q, and only looking through to the terminal outcome, which the dynamic planner does, restores the agreement.

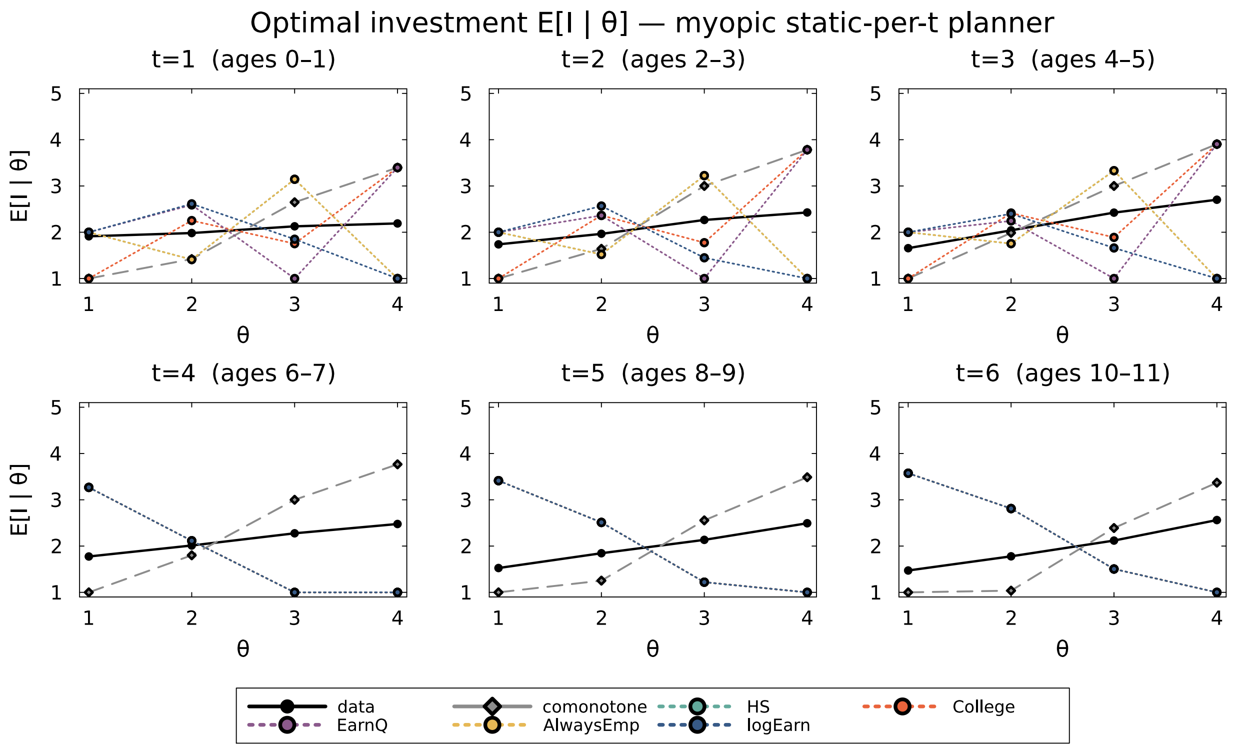

Figure 9: Optimal investment $E[I \mid \theta]$ under the myopic static-per-t planner, by developmental period and anchor. Contrast with the dynamic planner of Figure 5: from age six the two coincide, but in early childhood the myopic anchors spread apart, whereas the dynamic anchors remain nearly coincident.

C Optimal transfers

The final experiment is the most realistic. Each household keeps its own income y, and a piecewise-linear transfer $t(y)$ with non-negative values at the income knots \$5k/\$15k/\$40k/\$100k/\$300k is added before the behavioural CCP decides investment, subject to a per-child budget of

$10,000:

$$\max_{t_1,\dots,t_5 \geq 0} \; E[Q(\theta_T)] \quad \text{s.t.} \quad E[t(y)] \leq \$10{,}000,$$

where the constraint is the population-average transfer. Solving this for each of the five anchors (Figure 10), the optimal schedule is progressive but *non-monotone* in every case—transfers peak at the $15k-income knot rather than at the poorest $5k knot and fall to zero above $40k, so the budget is spent almost entirely on low- but not the lowest-income households. The welfare gains are $+1.0% (HS), +7.8% (college), +3.3% (earnings quintile), +1.8% (employment), and +0.4% (log earnings); college, whose value concentrates at the top of the skill distribution, is by far the most responsive to lifting low-income children into investment. The anchors differ only in how sharply they concentrate: maximizing high-school completion or employment targets the very poorest (essentially nothing above the $15k knot), whereas college spreads transfers a little further up—placing its largest above-bottom share at the $40k knot—with the earnings anchors in between.

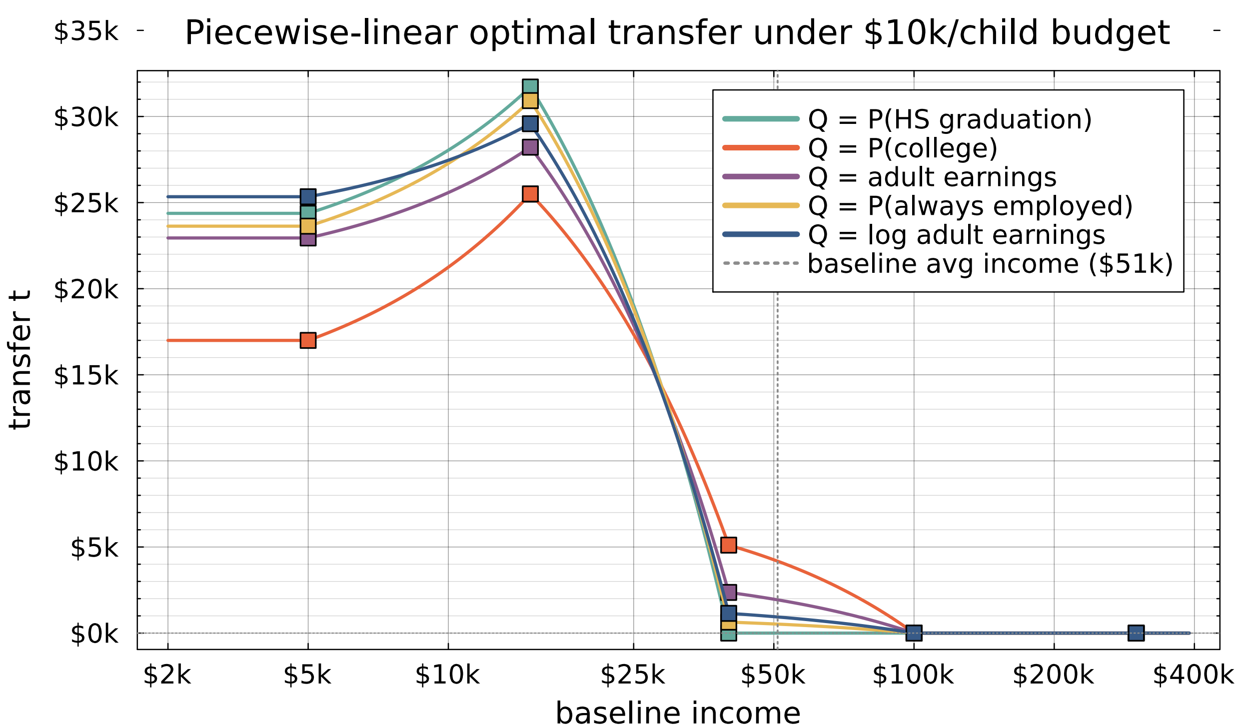

Figure 10: Optimal $10k-per-child transfer schedules, by anchor. Every schedule is progressive but non-monotone—transfers peak at the $15k-income knot (not the poorest $5k knot) and vanish above $40k. College spreads transfers furthest up the income distribution; high-school completion and employment target the poorest most sharply. Squares mark knot values.

D The Dynamic Social Planner Problem in its recursive form

Recall that the Social Planner Problem we study here is given by:

$$\hat{v}_0(F_{\theta_0}) = \sup_{\{F_{I_t|\theta_t}\}_{t=0}^{T}} \sum_{t=0}^{T+1} \int_{\theta_t} Q_t(c_t) dF_{\theta_t}(c_t)$$
$$\int_{I_t} \int_{\theta_t} b_t(i_t) dF_{I_t|\theta_t}(i_t|c_t) dF_{\theta_t}(c_t) \leq B_t \text{ for all } t$$
$$F_{\theta_t}(c_t) = \int_{\theta_{t-1}} \int_{I_{t-1}} F_{\theta_t|\theta_{t-1},I_{t-1}}(c_t|c_{t-1}, i_{t-1}) dF_{I_{t-1}|\theta_{t-1}}(i_{t-1}) dF_{\theta_{t-1}}(c_{t-1}) \text{ for all } t$$

In general, we can define the continuation value as:

$$\hat{v}_\tau(F_{\theta_\tau}) = \sup_{\{F_{I_t|\theta_t}\}_{t=\tau}^{T}} \sum_{t=\tau}^{T+1} \int_{\theta_t} Q_t(c_t) dF_{\theta_t}(c_t)$$
$$\int_{I_t} \int_{\theta_t} b_t(i_t) dF_{I_t|\theta_t}(i_t|c_t) dF_{\theta_t}(c_t) \leq B_t \text{ for all } t$$
$$F_{\theta_t}(c_t) = \int_{\theta_{t-1}} \int_{I_{t-1}} F_{\theta_t|\theta_{t-1},I_{t-1}}(c_t|c_{t-1}, i_{t-1}) dF_{I_{t-1}|\theta_{t-1}}(i_{t-1}) dF_{\theta_{t-1}}(c_{t-1}) \text{ for all } t$$

It will be useful to consider the recursive version of this problem.

The recursive social planner problem is given by:

$$v_t(F_{\theta_t}) = \sup_{F_{I_t|\theta_t}} \mathbb{E}_{F_{\theta_t}}[Q_t(\theta_t)] + v_{t+1}(F_{\theta_{t+1}}) \tag{Recursive SP}$$
$$\text{s.t} \int_{I_t} \int_{\theta_t} b_t(i) dF_{I_t|\theta_t}(i|c) dF_{\theta_t}(c) \leq B_t$$
$$F_{\theta_{t+1}}(z) = \int_{\theta_t} \int_{I_t} F_{\theta_{t+1}|\theta_t,I_t}(z|c,i) dF_{I_t|\theta_t}(i|c) dF_{\theta_t}(c)$$

with terminal condition:

$$v_{T+1}(F_{\theta_{T+1}}) = \mathbb{E}_{F_{\theta_{T+1}}}[Q_{T+1}(\theta_{T+1})]$$

The following proposition establishes that the recursive Social Planner Problem and the Sequential Social Planner Problem are equivalent:

Proposition 10. *For each $\tau = 0, \ldots, T+1$ the following equality holds:*

$$\hat{v}_t(F_{\theta_t}) = v_t(F_{\theta_t})$$

Proof. The proof proceeds by backward induction:

- **Base case $t = T+1$:** This case follows from the terminal condition in the recursive problem and the definition of the continuation value in the sequential problem for $\tau = T+1$

$$\hat{v}_{T+1}(F_{\theta_{T+1}}) = \mathbb{E}_{F_{\theta_{T+1}}}[Q_{T+1}(\theta_{T+1})] = v_{T+1}(F_{\theta_{T+1}})$$

- **Induction step** The induction hypothesis is:

$$\hat{v}_\tau(F_{\theta_\tau}) = v_\tau(F_{\theta_\tau}) \text{ for } \tau = t+1, \ldots, T+1$$

We want to show that:

$$\hat{v}_t(F_{\theta_t}) = v_t(F_{\theta_t})$$

For compactness of notation, $\mathcal{F}_{t+1}(F_{\theta_t})$ denote the set of feasible distributions $F_{\theta_{t+1}}$, and note that choosing such distribution is equivalent to choosing $F_{I_t|\theta_t}$. Note that

$$\begin{aligned}
v_t(F_{\theta_t}) &= \sup_{F_{\theta_{t+1}} \in \mathcal{F}_{t+1}(F_{\theta_t})} \left\{ \mathbb{E}_{F_{\theta_t}}[Q_t(\theta_t)] + v_{t+1}(F_{\theta_{t+1}}) \right\} \\
&= \sup_{F_{\theta_{t+1}} \in \mathcal{F}_{t+1}(F_{\theta_t})} \left\{ \mathbb{E}_{F_{\theta_t}}[Q_t(\theta_t)] + \hat{v}_{t+1}(F_{\theta_{t+1}}) \right\} \\
&= \sup_{F_{\theta_{t+1}} \in \mathcal{F}_{t+1}(F_{\theta_t})} \left\{ \int_{\theta_t} Q_t(c_t)\, dF_{\theta_t}(c_t) + \sup_{\{F_{\theta_{\tau+1}} \in \mathcal{F}_{\tau+1}(F_{\theta_\tau})\}_{\tau=t+1}^T} \sum_{\tau=t+1}^{T+1} \int_{\theta_\tau} Q_\tau(c_\tau)\, dF_{\theta_\tau}(c_\tau) \right\} \\
&= \int_{\theta_t} Q_t(c_t)\, dF_{\theta_t}(c_t) + \sup_{F_{\theta_{t+1}} \in \mathcal{F}_{t+1}(F_{\theta_t})} \sup_{\{F_{\theta_{\tau+1}} \in \mathcal{F}_{\tau+1}(F_{\theta_\tau})\}_{\tau=t+1}^T} \sum_{\tau=t+1}^{T+1} \int_{\theta_\tau} Q_\tau(c_\tau)\, dF_{\theta_\tau}(c_\tau) \\
&= \int_{\theta_t} Q_t(c_t)\, dF_{\theta_t}(c_t) + \sup_{\{F_{\theta_{\tau+1}} \in \mathcal{F}_{\tau+1}(F_{\theta_\tau})\}_{\tau=t}^T} \sum_{\tau=t+1}^{T+1} \int_{\theta_\tau} Q_\tau(c_\tau)\, dF_{\theta_\tau}(c_\tau) \\
&= \sup_{\{F_{\theta_{\tau+1}} \in \mathcal{F}_{\tau+1}(F_{\theta_\tau})\}_{\tau=t}^T} \sum_{\tau=t}^{T+1} \int_{\theta_\tau} Q_\tau(c_\tau)\, dF_{\theta_\tau}(c_\tau) \;=\; \hat{v}_t(F_{\theta_t}),
\end{aligned}$$

where the second equality follows from the Induction Hypothesis, the third equality follows from the definition of $\hat{v}_{t+1}$, the fourth equality follows from the fact that the first term in the sum does not depend on $F_{\theta_{t+1}}$, the fifth equality follows from $\sup_{x\in X}\sup_{y\in Y(x)} f(x,y) = \sup_{\{(x,y)\,:\,x\in X,\,y\in Y(x)\}} f(x,y)$ and the last equality follows from taking the constant term with respect to $F_{\theta_{t+1}}$ inside the supremum. Hence:

$$v_t(F_{\theta_t}) = \hat{v}_t(F_{\theta_t})$$

as desired.

$\square$

Before proving that the solution to this problem is monotonic in skills (that is, the optimal $F_{I_t|\theta_t}$ is increasing in a stochastic dominance sense in θ_t), it is useful to show that $v_t(F_{\theta_t})$ is monotonic in F_{θ_t}:

Lemma 2. *Suppose that for all t $F_{\theta_{t+1}|\theta_t,I_t}(z|c,i)$ is monotonic in c and Q_t is non-decreasing. Then F_{θ_t} FOSD F'_{θ_t} implies $v(F_{\theta_t}) \geq v(F'_{\theta_t})$*

Proof. The proof proceeds by backward induction:

- **Base case** $t = T+1$. Suppose F_{θ_t} FOSD F'_{θ_t}. Then, since Q_{T+1} is non-decreasing we have:

$$v(F_{\theta_t}) = \mathbb{E}_{F_{\theta_{T+1}}}[Q_{T+1}(\theta_{T+1})] \geq \mathbb{E}_{F'_{\theta_{T+1}}}[Q_{T+1}(\theta_{T+1})] = v_{T+1}(F'_{\theta_{T+1}})$$

- **Induction step:** Suppose that for $\tau = t+1,\dots,T+1$, F_{θ_τ} FOSD F'_{θ_τ} implies $v_\tau(F_{\theta_\tau}) \geq v_\tau(F'_{\theta_\tau})$. We want to show that F_{θ_t} FOSD F'_{θ_t} implies $v_t(F_{\theta_t}) \geq v_t(F'_{\theta_t})$

To show that, suppose that $F^*_{I_t|\theta_t}$ is the solution associated to F'_{θ_t}, and let $F^*_{I_t}$ and C^* be the marginal for investment and the copula associated to that solution, that is:

$$\begin{aligned} F^*_{I_t}(i) &= \int_{\theta_t} F^*_{I_t|\theta_t}(i|c)\, dF'_{\theta_t}(c) \\ C^*(u_1,u_2) &= \int_0^{u_1} F^*_{I_t|\theta_t}\Big((F^*_I)^{-1}(u_2)\,\Big|\,(F'_{\theta_t})^{-1}(s)\Big)\, ds \end{aligned}$$

Choosing the same marginal and the same copula under F_{θ_t} yields a marginal distribution of θ_{t+1} given by:

$$F_{\theta_{t+1}}(z) = \int_0^1\int_0^1 F_{t+1}(z|F^{-1}_{\theta_t}(u_1),(F^*_{I_t})^{-1}(u_2))dC^*(u_1,u_2)$$

Now, since F_{θ_t} FOSD F'_{θ_t} we have that

$$F_{\theta_t}^{-1}(u_1) \geq (F'_{\theta_t})^{-1}(u_1) \text{ for all } u_1 \in [0,1]$$

This and the monotonicity of $F_{t+1}(\cdot|c,i)$ in c implies:

$$F_{\theta_{t+1}}(z) = \int_0^1 \int_0^1 F_{t+1}(z|F_{\theta_t}^{-1}(u_1), (F_{I_t}^*)^{-1}(u_2)) dC^*(u_1,u_2) \leq$$
$$\int_0^1 \int_0^1 F_{t+1}(z|(F'_{\theta_t})^{-1}(u_1), (F_{I_t}^*)^{-1}(u_2)) dC^*(u_1,u_2) = F'_{\theta_{t+1}}(z)$$

where $F'_{\theta_{t+1}}$ is the marginal distribution of θ_{t+1} implied by F'_{θ_t} and $F^*_{I_t|\theta_t}$. Since z is arbitrary $F_{\theta_{t+1}}$ FOSD dominates $F'_{\theta_{t+1}}$.

Hence, we have:

$$v_t(F_{\theta_t}) \geq \mathbb{E}_{F_{\theta_t}}[Q_t(\theta_t)] + v_{t+1}(F_{\theta_{t+1}}) \geq \mathbb{E}_{F'_{\theta_t}}[Q_t(\theta_t)] + v_{t+1}(F'_{\theta_{t+1}}) = v_t(F'_{\theta_t})$$

where the first inequality follows from the fact that choosing the same marginal and the same copula for F_{θ_t} as in F'_{θ_t} is feasible, the second inequality follows from the fact that $\mathbb{E}_{F_{\theta_t}}[Q_t(\theta_t)] \geq \mathbb{E}_{F'_{\theta_t}}[Q_t(\theta_t)]$, $F_{\theta_{t+1}}$ FOSD dominates $F'_{\theta_{t+1}}$ and the induction hypothesis, and the last equality follows from the fact that $F^*_{I_t|\theta_t}$ is optimal under F'_{θ_t}. □

Now we can establish our co-monotonicity result for the recursive social planner problem:

Proposition 11. *Suppose $F_{t+1}(x|\theta_t, I_t)$ is ordinally complementary for all $t = 0, \ldots, T$ and that for all $y = (\theta_t, I_t), y' = (\theta'_t, I'_t)$ with $y \geq y'$*

$$F_{t+1}(x|y) \text{ FOSD } F_{t+1}(x|y')$$

and suppose that each Q_t is non-decreasing and the recursive social planner problem has a solution at time t for initial distribution F_{θ_t}. Then it has a co-monotone solution

Proof. Suppose the social planner problem has a solution $F^*_{I_t|\theta_t}$ with marginal $F^*_{I_t}$. From Corollary 1, the co-monotone solution of the second-stage optimal transport problem yields a distribution that FOSD the one implied by this solution. By Lemma 2, it yields a higher future value. Since the flow value is equal for both (because it only depends on F_{θ_t}), this completes the proof. □

E Monotonicity in the optimal tracking problem

Section 3.2 showed that ordinal complementarity delivers a comonotone solution to the sequential planner problem (6). This appendix treats a second dynamic problem, in which the planner observes only initial skill and commits to a whole sequence of investments, and shows that ordinal complementarity again delivers monotonicity—though here a strict form of the property is required, for reasons explained below.

E.1 The problem and the result

The planner observes only the initial level of skills and assigns children to sequences of investment $\{I_t(\theta_0)\}_{t=0}^T$:

$$\sup \int_{\Theta_0} \mathbb{E}[Q(\theta_{T+1})|\theta_0, I_0(\theta_0), \dots, I_T(\theta_0)] \quad \text{s.t } \{I_t()\}_t^T \in \mathbb{I} \tag{13}$$

where Θ_0 is a compact subset of $\mathbb{R}$ and here $\mathbb{I}$ is a subset of the set of bounded, measurable functions from Θ_0 to $\mathcal{I}$, where $\mathcal{I}$ is a compact subset of $\mathbb{R}^T$.

We can think of this problem as reflecting educational tracking, which might be justified if the cost of re-assigning children to tracks each period is large.

In order to prove our monotonicity result for this problem we need that the feasibility of a given investment policy only depends on the marginal distribution of investment that it induces in each period. The following assumption formalizes that:

Assumption 2 (Rearrangement invariance)**.** The constraint set $\mathbb{I}$ is closed under equimeasurable rearrangement: if $I = (I_0, \dots, I_T) \in \mathbb{I}$ and $\tilde{I} = (\tilde{I}_0, \dots, \tilde{I}_T)$ satisfies $\tilde{I}_t \sharp \mu_{\theta,0} = I_t \sharp \mu_{\theta,0}$ for every t, then $\tilde{I} \in \mathbb{I}$.

Note that the constraint set for this Social Planner Problem is very general, and it allows for, for example, period-by-period budget constraints (no borrowing) and for intertemporal budget constraints.

Proposition 12 (Dynamic comonotonicity in tracking)**.** *Suppose each one-step technology $F_{t+1}(\cdot \mid \theta_t, I_t)$ is strictly ordinally complementary in the sense of Definition 6, is FOSD-monotone in (θ_t, I_t) and satisfies (S1)–(S2) below, the Markov Kernel induced by it satisfies the Weak Feller Property, and F_{θ_0} has no mass points. Moreover, assume that 2 holds. Then, for an arbitrary strictly increasing, continuous Q, if a solution to the optimal tracking* *problem* (13) *exists, it has a non-decreasing solution, i.e.* $I_t^*(\theta_0)$ *is non-decreasing for every* t.

The proof of Proposition 12 uses Theorem 4.1 of Carlier (2003), which delivers a monotone optimal transport solution when the cost is *strictly* monotone of order 2, which is a property implied by strict submodularity (so for our purpose here we will require strict supermodularity, since we are considering a maximization problem instead of a minimization problem, like Carlier (2003) does). That is why this appendix develops strict counterparts of Definition 1 and Proposition 2 before turning to the sequential results.

E.2 Strict ordinal complementarity

Supermodularity has no content at pairs that are ordered in the first two coordinates, since then $\{x \wedge x', x \vee x'\} = \{x, x'\}$ and both sides of (4) coincide identically. Such pairs impose no restriction in Definition 1 but admit no strict inequality, which is why the strict notion is stated over the following set.

Definition 5 (Crossing pairs)**.** A pair $(x, x') \in \mathcal{X} \times \mathcal{X}$ is a *crossing pair* in the first two coordinates if it agrees off those coordinates and orders them in opposite directions:

$$\mathcal{P} \;=\; \Big\{(x,x') \in \mathcal{X} \times \mathcal{X} \;:\; x_j = x'_j \text{ for all } j > 2 \;\text{ and }\; (x_1 - x'_1)(x_2 - x'_2) < 0\Big\}.$$

Definition 6 (Strict ordinal complementarity)**.** X_1 and X_2 are *strict* ordinal complements in Y if they are ordinal complements in the sense of Definition 1 and, in addition, for every crossing pair $(x, x') \in \mathcal{P}$ the inequality (4) is strict for at least one $y \in \mathcal{Y}$.

Strictness is required at some y rather than at every y. The two sides of (4) necessarily coincide at the top of $\mathcal{Y}$, so a pointwise requirement could never be met when $\mathcal{Y}$ has a maximum—as it does in the discrete setting of Sections 4–6—and strictness at a single y is exactly what a strictly increasing valuation detects, which is what makes Proposition 13 an equivalence rather than a one-way implication.

Proposition 13 (Strict supermodularity of expectations)**.** *X_1 and X_2 are strict ordinal complements in Y if and only if, for every measurable, strictly increasing, integrable Q and every crossing pair $(x, x') \in \mathcal{P}$,*

$$E[Q(Y) \mid x \wedge x'] + E[Q(Y) \mid x \vee x'] > E[Q(Y) \mid x] + E[Q(Y) \mid x']. \tag{14}$$

The proof rests on the following strengthening of the stochastic-dominance characterization used in the proof of Proposition 2, recorded here for completeness.

Lemma 3 (Strict stochastic dominance)**.** *Let μ and ν be probability measures on $\mathbb{R}$ with CDFs F_μ and F_ν. Then $F_\mu(y) \le F_\nu(y)$ for all y, with strict inequality for at least one y, if and only if $\int Q\, d\mu > \int Q\, d\nu$ for every measurable, strictly increasing Q with $\int |Q|\, d\mu < \infty$ and $\int |Q|\, d\nu < \infty$.*

Proof. (*Only if.*) Let $U \sim \mu$ and $V \sim \nu$, so that U first-order stochastically dominates V, and let Q be as in the statement, so that $E[Q(U)]$ and $E[Q(V)]$ are finite. Monotone transformations preserve first-order stochastic dominance (Shaked and Shanthikumar, 2007, §1.A.3), so $Q(U)$ dominates $Q(V)$. Suppose $E[Q(U)] = E[Q(V)]$. Two random variables so ordered with equal finite means are equal in distribution (Shaked and Shanthikumar, 2007, Thm. 1.A.8), so $Q(U)$ and $Q(V)$ have the same law. Since Q is strictly increasing, $Q(U) \le Q(y)$ if and only if $U \le y$, whence $F_\mu(y) = \mathbb{P}(Q(U) \le Q(y)) = \mathbb{P}(Q(V) \le Q(y)) = F_\nu(y)$ for every $y \in \mathbb{R}$, contradicting $F_\mu \ne F_\nu$. Therefore $E[Q(U)] \ne E[Q(V)]$, and since dominance and finiteness of the means give $E[Q(U)] \ge E[Q(V)]$, the inequality is strict.

(*If.*) The function arctan is strictly increasing, continuous and bounded by $\pi/2$, hence measurable and integrable under any probability measure, so the hypothesis gives $\int \arctan d\mu > \int \arctan d\nu$ and in particular $F_\mu \ne F_\nu$. Suppose next that $\Delta := F_\mu(\bar{y}) - F_\nu(\bar{y}) > 0$ for some $\bar{y}$, and for $\varepsilon > 0$ put $Q_\varepsilon(t) = \mathbf{1}\{t > \bar{y}\} + \varepsilon \arctan t$. Then Q_ε is measurable, being the sum of the indicator of a Borel set and a continuous function; it is strictly increasing, since the step term is non-decreasing and $\varepsilon \arctan$ is strictly increasing; and it is bounded by $1+\varepsilon\pi/2$, hence integrable under both measures. Writing $A = \int \arctan d\mu - \int \arctan d\nu$ and using $\int \mathbf{1}\{t > \bar{y}\}\, d\mu = 1 - F_\mu(\bar{y})$,

$$\int Q_\varepsilon\, d\mu - \int Q_\varepsilon\, d\nu = \Big[(1 - F_\mu(\bar{y})) - (1 - F_\nu(\bar{y}))\Big] + \varepsilon A = -\Delta + \varepsilon A.$$

Both integrals defining A lie in $[-\pi/2, \pi/2]$, so $|A| \le \pi$; choosing $\varepsilon \in (0, \Delta/\pi)$ makes the display at most $-\Delta + \varepsilon\pi < 0$, a contradiction. Hence $F_\mu \le F_\nu$ pointwise, and with $F_\mu \ne F_\nu$ the proof is complete. □

Proof of Proposition 13. Fix a crossing pair $(x, x') \in \mathcal{P}$ and let μ and ν be the measures of the previous proof. Definition 6 states that $F_\mu(y) \le F_\nu(y)$ for all $y \in \mathcal{Y}$, and strictness states that the inequality is strict for at least one y; (14) states, after multiplying through by $\frac{1}{2}$, that $\int Q\, d\mu > \int Q\, d\nu$. Lemma 3 gives the equivalence of these two statements at this pair. Since $(x, x') \in \mathcal{P}$ was arbitrary and both conditions quantify over crossing pairs, the proposition follows. $\square$

E.3 Sequential complementarity

The remaining results carry ordinal complementarity from the one-step technologies to the composite technology mapping $(\theta_0, I_0, \dots, I_T)$ into terminal skill. Throughout, the one-step conditionals have no mass points, so the integration by parts is exact. Each result has a weak form and a strict form; the strict forms rest on the following strengthening of the primitives. For all t:

(S1) $\partial_\theta \overline{F}_{Y|X}(x \mid \theta, I)$ exists, is continuous in (θ, I), and satisfies $\partial_\theta \overline{F}_{Y|X}(x \mid \theta, I) > 0$ for x and θ in the interiors of $\mathcal{Y}$ and Θ *(strict dominance in skill)*;

(S2) $\overline{F}_{Y|X}(x \mid \theta, I) > \overline{F}_{Y|X}(x \mid \theta, I')$ whenever $I > I'$ and x is interior *(strict dominance in investment)*;

(S3) $F_{t+1}(\cdot \mid \theta_t, I_t)$ is a strict ordinal complement in the sense of Definition 6.

Lemma 4 (Sequential monotonicity)**.** *Suppose that for all t, $F_{t+1}(\cdot \mid \theta_t, I_t)$ first-order stochastically dominates $F_{t+1}(\cdot \mid \theta'_t, I'_t)$ whenever $(\theta_t, I_t) \geq (\theta'_t, I'_t)$. Then for any predetermined investment paths $I^t \geq (I^t)'$, $\theta_0 \geq \theta'_0$, $F_{t+1}(\cdot \mid \theta_0, I^t)$ FOSD $F_{t+1}(\cdot \mid \theta'_0, (I^t)')$. If in addition (S1) and (S2) hold and $(\theta_0, I^t) \neq (\theta'_0, (I^t)')$, the dominance is strict on the interior:*

$$F_{t+1}(\theta \mid \theta_0, I^t) \; < \; F_{t+1}\Big(\theta \mid \theta'_0, (I^t)'\Big) \qquad \textit{for every } \theta \textit{ in the interior of } \Theta.$$

Proof of Lemma 4. By induction on t. For $t = 0$ the claim is the maintained property that $F_{t+1}(\cdot \mid \theta_t, I_t)$ first-order stochastically dominates $F_{t+1}(\cdot \mid \theta'_t, I'_t)$ when $(\theta_t, I_t) \geq (\theta'_t, I'_t)$. For the step, the law of total probability gives

$$\overline{F}_{t+1}(x \mid \theta_0, I^t) = \int \overline{F}_{Y|X}(x \mid \theta_t, I_t)\, dF(\theta_t \mid \theta_0, I^{t-1}).$$

Take predetermined paths $I^t \geq (I^t)'$, so $I_t \geq I'_t$ and $I^{t-1} \geq (I^{t-1})'$. Raising I_t raises $\overline{F}_{Y|X}(x \mid \theta_t, I_t)$ pointwise in θ_t (dominance in investment). Raising I^{t-1} makes $F(\cdot \mid \theta_0, I^{t-1})$ stochastically larger by the induction hypothesis; since $\theta_t \mapsto \overline{F}_{Y|X}(x \mid \theta_t, I_t)$ is non-decreasing (dominance in skill), integrating it against a stochastically larger distribution of θ_t raises the integral. Both effects increase $\overline{F}_{t+1}(x \mid \theta_0, I^t)$, so $\overline{F}_{t+1}(x \mid \theta_0, I^t) \geq \overline{F}_{t+1}(x \mid \theta_0, (I^t)')$ for every x, which is the asserted dominance.

For the strict claim, assume (S1)–(S2) and run the same induction, tracking strictness. At $t = 0$ the claim is (S1) if $\theta_0 > \theta_0'$ and (S2) if $I_0 > I_0'$. At the step, suppose first $I_t > I_t'$ with the remaining coordinates fixed: (S2) raises the integrand strictly and pointwise in θ_t, so the integral rises strictly. Suppose instead $I_t = I_t'$ and (θ_0, I^{t-1}) is raised: the induction hypothesis makes $F(\cdot \mid \theta_0, I^{t-1})$ strictly dominated on the interior, and (S1) makes $\theta_t \mapsto \overline{F}_{Y|X}(x \mid \theta_t, I_t)$ strictly increasing and bounded, so Lemma 3 gives a strict increase in the integral. A general increase composes the two cases and at least one is strict. Passing from $\overline{F}_{t+1}$ to F_{t+1} reverses the inequality. □

Proposition 14 (Sequential complementarity)**.** *Suppose $F_{t+1}(\cdot \mid \theta_t, I_t)$ is ordinally complementary for all t and FOSD-monotone in (θ_t, I_t). Then $\overline{F}_{t+1}(x \mid \theta_0, I_0, \dots, I_T)$ is supermodular in $(\theta_0, I_0, \dots, I_T)$. If in addition (S1)–(S3) hold, then for every crossing pair in $(\theta_0, I_0, \dots, I_T)$ the supermodularity inequality is strict for at least one $x \in \mathcal{Y}$; equivalently, $F_{t+1}(\cdot \mid \theta_0, I_0, \dots, I_T)$ is a strict ordinal complement in every pair of its conditioning coordinates.*

Proof of Proposition 14. By induction on t. For $t = 0$ the claim is ordinal complementarity of $F_{t+1}(\cdot \mid \theta_0, I_0)$. Assume $\overline{F}_t(\cdot \mid \theta_0, I^{t-1})$ is supermodular in (θ_0, I^{t-1}). By the law of total probability and integration by parts—the conditionals have no mass points, which removes the lower boundary term—

$$\overline{F}_{t+1}(x \mid \theta_0, I^t) = \int \overline{F}_{Y|X}(x \mid \theta, I_t)\, dF(\theta \mid \theta_0, I^{t-1}) = \overline{F}_{Y|X}(x \mid \bar{\theta}, I_t) - \int \partial_\theta \overline{F}_{Y|X}(x \mid \theta, I_t)\, F(\theta \mid \theta_0, I^{t-1})\, d\theta.$$

Take $I_t \geq I_t'$ without loss. The boundary terms $\overline{F}_{Y|X}(x \mid \bar{\theta}, \cdot)$ cancel in the mixed second difference, so supermodularity of $\overline{F}_{t+1}$ in (θ_0, I^t) is equivalent to $\int (aA - bB)\, d\theta \geq 0$, where

$$\begin{gathered} a(\theta) = \partial_\theta \overline{F}_{Y|X}(x \mid \theta, I_t), \qquad b(\theta) = \partial_\theta \overline{F}_{Y|X}(x \mid \theta, I_t') \\ A(\theta) = F(\theta \mid \theta_0, I^{t-1}) - F\Big(\theta \mid \max\{\theta_0, \theta_0'\}, I^{t-1} \vee (I^{t-1})'\Big), \\ B(\theta) = F\Big(\theta \mid \min\{\theta_0, \theta_0'\}, I^{t-1} \wedge (I^{t-1})'\Big) - F\Big(\theta \mid \theta_0', (I^{t-1})'\Big) \end{gathered}$$

Two facts close it, pointwise in θ. First, $a \geq b \geq 0$: ordinal complementarity makes $\partial_\theta \overline{F}_{Y|X}$ non-decreasing in I, and dominance in skill makes $\overline{F}_{Y|X}$ non-decreasing in θ. Second, the induction hypothesis—$\overline{F}_t$ supermodular, equivalently F_t submodular, in (θ_0, I^{t-1})—gives $A \geq B$, and dominance in the predetermined path gives $B \geq 0$. Since

$$aA - bB \;=\; (a - b)\, A \;+\; b\,(A - B), \tag{15}$$

both terms are non-negative, the larger derivative multiplies the larger bracket, and the inequality holds for every x. This is supermodularity of $\overline{F}_{t+1}$ in (θ_0, I^t), completing the induction.

For the strict claim, assume (S1)–(S3) and strengthen the induction hypothesis to: for every crossing pair in (θ_0, I^{t-1}) there is an x and a θ at which the supermodularity inequality for $\overline{F}_t$ is strict. At $t = 0$ this is (S3). At the step, fix a crossing pair in (θ_0, I^t); which term of (15) delivers strictness depends on the pair.

Pairs involving I_t. Let the crossing pair be (I_t, c) with c one of $\theta_0, I_0, \ldots, I_{t-1}$, so $I_t > I'_t$ and the two predetermined paths differ only in c. They are then ordered, so $A = B$ and the second term of (15) vanishes; strictness must come from $(a - b)A$. The two paths in the definition of A differ strictly in c, so Lemma 4 gives $A(\theta) > 0$ at every interior θ. And (S3), applied at the crossing pair $\big((\theta', I'_t), (\theta, I_t)\big)$ with $\theta > \theta'$, supplies an x with $\int_{\theta'}^{\theta} [a(u) - b(u)]\, du > 0$; since $a - b$ is continuous by (S1), $a > b$ on a subinterval of positive Lebesgue measure. At that x the integrand $(a - b)A$ is non-negative everywhere and strictly positive on a set of positive measure.

Pairs not involving I_t. Let the crossing pair lie in (θ_0, I^{t-1}), so $I_t = I'_t$, hence $a = b$ and the first term of (15) vanishes; strictness must come from $b\,(A - B)$. The induction hypothesis supplies an x and a θ_* with $A(\theta_*) > B(\theta_*)$. Since $A - B$ is a finite linear combination of conditional CDFs in θ, it is right-continuous, so $A > B$ on $[\theta_*, \theta_* + \delta)$ for some $\delta > 0$, which may be taken small enough that the interval is interior; there $b(\theta) > 0$ by (S1). Again the integrand is non-negative everywhere and strictly positive on a set of positive measure.

In both cases $\int (aA - bB)\, d\theta > 0$ at some x, so the supermodularity inequality for $\overline{F}_{t+1}(\cdot \mid \theta_0, I^t)$ is strict at that x, which completes the induction.

The argument is written for conditionals without mass points, so the integration by parts is exact. For the discrete estimated technology of Section 5 the same conclusion holds with sums replacing integrals and the cell difference-in-differences of Section 4 replacing $\partial_\theta \overline{F}_{Y|X}$; the algebra is identical, and the right-continuity step is immediate, since a strict inequality at a cell is already strict on a set of positive counting measure. □

Remark 1. Both terms of (15) are needed, so (S1)–(S3) are not redundant. Strict ordinal complementarity of the one-step technologies alone does not suffice: it delivers $a > b$, but $(a - b)A$ vanishes wherever A does, and at pairs not involving I_t the factor $a - b$ is identically zero. Symmetrically, strict dominance alone delivers $b > 0$ and $A > 0$ but leaves $a = b$ and $A = B$ at the pairs where strictness must originate.

E.4 Proof of Proposition 12

Proof of Proposition 12. By assumption, a Borel-measurable solution $\tilde{I} = (\tilde{I}_0, \tilde{I}_1, \dots, \tilde{I}_T)$ exists. We want to show that a non-decreasing solution $I^* = (I^*_0, I^*_1, \dots, I^*_T)$ also exists. Let $\mu_{I,t} := \tilde{I}_t \sharp \mu_{\theta,0}$ be the push-forward measure of $\mu_{\theta,0}$ through $\tilde{I}_t$, where $\mu_{\theta,0}$ is the measure associated to F_{θ_0}. Moreover, note that F_{θ_0} has no mass points, Θ, Θ_0 and $\mathcal{I}$ are compact, and $\mathbb{E}[Q(\theta_{T+1})|\theta_0, I_0, \dots, I_T]$ is continuous and bounded [10]. Moreover, $\mathbb{E}[Q(\theta_{T+1})|\theta_0, I_0, \dots, I_T]$ is strictly supermodular by Propositions 13 and 14, so -$\mathbb{E}[Q(\theta_{T+1})|\theta_0, I_0, \dots, I_T]$ is strictly submodular, which implies monotonicity of order 2 in the language of Carlier (2003). Hence, by Theorem 4.1 in Carlier (2003) there exist a non-decreasing $I^* = (I^*_0, I^*_1, \dots, I^*_T)$ solution to:

$$\inf_{(I_0,\dots,I_T)\in\Gamma} - \int_{\Theta_0} \mathbb{E}[Q(\theta_{T+1})|\theta_0, I_0(\theta_0), \dots, I_T(\theta_0)] dF_{\theta_0}(\theta_0) \tag{16}$$

where

$$\Gamma := \Big\{ s = (s_0, \dots, s_T) : \quad \text{each } s_t \colon \mathbb{R} \to \mathbb{R} \text{ is Borel, and } s_t \sharp \mu_{\theta,0} = \mu_{I,t} \text{ for } i = 0, \dots, T \Big\}.$$

Since $\tilde{I}$ is feasible in that set, I^* yields a weakly smaller objective in (16). Moreover, since I^*_t has the same marginal as $\tilde{I}_t$ for every t, it is also feasible in the optimal tracking problem by Assumption 2, and hence a solution. Since I^*_t is non-decreasing for all $t = 0, \dots, T$, this completes the proof. □

[10]Because Q is continuous, it maps compact sets to compact sets. Moreover, since $F_t(\theta_{t+1}|\theta_t, I_t)$ satisfies the Weak Feller Property for each t, so does $F_t(\theta_{t+1}|\theta_0, I_0, \dots, I_T)$, so $\mathbb{E}[Q(\theta_{T+1})|\theta_0, I_0, \dots, I_T]$ is continuous and bounded