Technological Change and Workforce Composition: Causal Evidence from R&D Subsidies in Digitalization

Annalisa Caloffi (University of Florence)
Giuseppe Caruso (University of Florence)
Martina Colucci (University of Florence)
Natalia Faraoni (IRPET)
Marco Mariani (IRPET, presenting author)
Alessandra Mattei (University of Florence)
Fiammetta Menchetti (University of Florence)

Florence, 23 October 2025

Digitalization, Automation and Workforce Composition: The Debate

A renewed question in economics and policy

- The spread of automation and digital technologies has reignited the debate on how technological change reshapes labor demand and skills (Autor 2015; Acemoglu & Restrepo 2018; Bessen 2019).
- Beyond job quantity, digital transformation affects the **composition of the workforce**:
  - Growing demand for ICT, data, and engineering skills (Brynjolfsson & McAfee 2014).
  - Decline of routine and manual tasks (Goos, Manning & Salomons 2014).
  - Hybrid profiles combining technical and organizational capabilities (Bessen 2022).
- Existing empirical work focuses on **automation** or on **digitalization**, but does not distinguish between latent **innovation strategies** (seeking efficiency vs. expansion) nor evaluate their causal effects on employment and productivity (Balsmeier & Woerter 2019; Domini et al. 2021, 2022)

Policy Context and Knowledge Gap

Why this study?

- European and regional innovation policies invest heavily in **digital R&D programs**, yet there is limited causal evidence on their effectiveness.
- Most evaluations do not distinguish between different **innovation strategies** — whether firms pursue efficiency-oriented or expansion-oriented digitalization — nor assess how policy effects vary across these strategies.

Our contribution

- First **causal evaluation** of a public R&D subsidy program for digitalization.
- Identification of latent **innovation strategies** (efficiency-seeking vs. expansion-seeking) through semantic analysis of project texts, allowing causal inference on **heterogeneous effects by strategy**.
- Integration of a **modern Regression Discontinuity Design** with **Principal Stratification** (Frangakis & Rubin, 2002; Li et al., 2015).
- The framework moves beyond descriptive or correlational approaches, linking **program design**, **firm behavior**, and **workforce outcomes**.

Key Research Questions

Do public R&D subsidies for digitalization increase productivity, reshape workforce composition, or both?

And how do these effects differ between efficiency-oriented and expansion-oriented innovation strategies?

At a Glance

- **Causal design:** Fuzzy Regression Discontinuity embedded in a Principal Stratification framework; local randomization within subpopulation U_{s0}.
- **Data:** Universe of 58k firms (Tuscany, 2012–2021), linked administrative and labor register data, plus project texts for participants.
- **Main finding:** *Workforce upgrading* — growth of specialized and high-skill roles — with limited direct effects on average productivity.
- **Heterogeneity:** innovation strategy matters — **efficiency-oriented** projects $\Rightarrow$ productivity gains; **expansion-oriented** projects $\Rightarrow$ employment growth.

Policy Program and Dataset

Policy: POR–FESR 2014–2020, Action 1.1.5 (R&D projects; firm-level and consortium tracks).

Eligibility rule:

- Firms with non-decreasing turnover in the pre-period qualify as *dynamic* (eligible).
- Non-dynamic firms may join through a *consortium* (majority of dynamic firms required).

Program participation (firms):

- Total firms in the database: **57,989**.
- **27,558** eligible (dynamic) and **30,431** non-eligible.
- Among eligible: **599** participants, of which **343** admitted to funding.
- Among non-eligible: **106** participants (in consortia), of which **60** admitted.

Data sources:

- **Firm accounts:** AIDA balance-sheet data (2012–2021) — value added, turnover, fixed assets.
- **Employment:** *Sistema Informativo Lavoro* (Regione Toscana, 2014–2022) — positions by qualification (ISTAT 2011, 1–8).
- **Program data:** POR–FESR 1.1.5 administrative records and project texts — participation, funding status, and semantic classification.

Semantic Classification of Projects

Objective: Identify heterogeneous innovation strategies within subsidized R&D projects.

Method: Text mining and manual validation of project descriptions (2014 call).

- Classification distinguishes between two **innovation orientations:**
  1. **Efficiency-oriented projects:** focus on process innovation and internal optimization (automation, integration, quality control).
  2. **Expansion-oriented projects:** focus on product innovation and market or scale expansion (new product lines, new clients, internationalization).
- Only projects approved under the 2014 call are used for classification (text availability criterion).

Distribution of Projects by Innovation Orientation

Universe: 2014 call, 170 approved projects (out of 221 total proposals).

Category	Total	Approved	Not approved
Efficiency-oriented (process)	40	30	10
Expansion-oriented (product)	159	121	38
Undetermined / Not classifiable	22	19	3

Key patterns:

- Expansion-oriented projects dominate the portfolio, reflecting a stronger focus on product development and scaling.
- Efficiency-oriented projects are fewer but tend to involve higher technological content and internal process reengineering.
- With an acceptance rate of about 75–76% in each category, expansion-oriented projects remain the majority among funded ones.
- These two groups define the heterogeneity dimension used in causal estimation (Section *Heterogeneity via Semantic Classification*).

Analytical Roadmap: Why This Design

1) Fuzzy RD: modern view

- The eligibility rule affects the *probability*, not the certainty, of treatment.
- Near the cutoff, assignment behaves as **as-if random** for a local population.
- $\Rightarrow$ **Gain: External validity** — effects generalize to a local population, not just a single cutoff point.

2) From assignment to observed and latent structure

- After assignment, we observe participation and funding $(Z, P, A) \in \{0, 1\}^3$, forming multiple **observed groups**.
- Beneath these lie **latent principal strata**, defined by firms' potential participation and subsidy receipt under each eligibility state.
- $\Rightarrow$ **Gain: Interpretability** — enables exploration of distinct **causal pathways**: **pure participation**, **pure subsidy**, and their **joint effect**.

3) Bayesian inference

- Observed groups are **mixtures of latent strata**; membership is unobserved.
- This setting is **weakly identified**: strata membership and outcomes are jointly inferred, not separately identified.
- Bayesian estimation coherently integrates uncertainty in both **strata membership** and **outcomes**, addressing weak identifiability through joint posterior inference (and flexible parametric assumptions).

Fuzzy RD Design as Local Experiment

Forcing variable: change in turnover (ΔS) over the pre-policy period.
Threshold at 0 defines "dynamic" (eligible) vs. "non-dynamic" (ineligible) firms.

Fuzzy nature of the design:

- Not all eligible firms participate or receive the subsidy.
- Some non-eligible firms access treatment via consortia.

Modern RD view: eligibility affects treatment only probabilistically, creating a **fuzzy RD**.
Within a local window around the threshold and conditional on rich pre-treatment covariates, eligibility status Z can be interpreted as *as-if randomly assigned*.

Selecting the Local Subpopulation U_{s0}

Goal: identify the subset of firms, U_{s0}, where the RD design behaves as a **locally randomized experiment**, conditional on covariates (local overlap + local unconfoundedness).

Selection procedure:

1. Estimate outcome-specific mixed-effect models on each side of the threshold, allowing firm-level slopes in ΔS (change in turnover).
2. Exclude firms showing significant slope differences ($p < 0.05$).
3. Define U_{s0} as the intersection of firms satisfying above condition for all outcomes.

Resulting sample: U_{s0} includes **2,006 firms** (795 eligible, 1,211 non-eligible).

Group	Total	Eligible	Non-eligible
Non-participants (NP)	1,315	209	1,106
Participants admitted (PA)	391	332	59
Participants not admitted (PNA)	300	254	46

Why Principal Stratification within a RD?

Classical RD interpretation:

- In the traditional view (Hahn et al., 2001), the **estimand is a causal effect at a threshold point**.
- Identification relies on smoothness of the conditional expectation around the cutoff.
- There is no **population of units** assigned to treatment or control — only a discontinuity in a continuous function.
- $\Rightarrow$ No meaningful way to define principal strata or potential outcomes for post-assignment variables.

Modern RD interpretation (Li et al., 2015; Cattaneo et al., 2015)

- Focuses on a **local population** U_{s0} of units near the threshold, where assignment Z behaves as *as-if random*.
- Within U_{s0}, the design becomes a **locally randomized experiment** with a binary assignment variable.
- This framework allows the use of **Principal Stratification** to decompose causal effects by post-assignment variables such as participation (P) and benefit receipt (A).

Key implication: In the classical RD there is a point; in the modern RD there is a population — and thus, a structure to stratify.

Principal Stratification: Setup

Post-assignment variables:

- P_i: participation in the program,
- A_i: receipt of the subsidy,
- Y_i: outcome variable.

Each firm i has potential values $\{P_i(0), P_i(1), A_i(0), A_i(1), Y_i(0), Y_i(1)\}$ depending on eligibility $Z \in \{0, 1\}$.

Principal strata: defined using $\{P_i(0), P_i(1), A_i(0), A_i(1)\}$, each taking values in $\{0, 1\}$. Out of $2^4 = 16$ resulting strata, 7 are impossible as they would imply subsidy without participation.

Monotonicity assumptions:

$$P_i(1) \geq P_i(0), \quad A_i(1) \geq A_i(0)$$

(no defiers). This rules out cases where becoming eligible decreases participation or benefit receipt.

These assumptions reduce the 9 remaining potential profiles to only **6** that are consistent with monotonicity.

(Full logical enumeration of strata available in Appendix.)

Principal Strata and Exclusion Restrictions

Stratum	Varies with Z	Label	Effect of Eligibility
$\bar{P}P\bar{A}\bar{A}$	P only	*Stimulated*	May stem from participation only.
$\bar{P}P\bar{A}A$	P, A	*Activated*	May stem from both participation and funding.
$PP\bar{A}A$	A only	*Reinforced*	May stem from funding among always-participants.
$\bar{P}\bar{P}\bar{A}\bar{A}$	none	*Unresponsive*	Impossible: lack of causal channel.
$PP\bar{A}\bar{A}$	none	*Aspirational*	Impossible: lack of causal channel.
$PPAA$	none	*Self-sufficient*	Impossible: lack of causal channel.

Table: Principal strata, variation with eligibility (Z), labels, and meaning of the eligibility effect.

Exclusion restrictions: For the last three strata, eligibility Z alters neither participation or benefit receipt, so any observed difference in outcomes cannot be causally attributed to the program. Hence, for these strata we impose:

$$Y_i(1) = Y_i(0)$$

and restrict estimation to the three "relevant" strata where Z changes P and/or A.

Estimands and Aggregation

Stratum-specific causal estimands:

$$\mathrm{PCE}_{g,t}^{(\ell)} = \mathbb{E}[Y_{it}^{(\ell)}(1) - Y_{it}^{(\ell)}(0) \mid G_i = g, i \in U_{s0}]$$

where $g \in \{\bar{P}P\bar{A}\bar{A},\ \bar{P}P\bar{A}A,\ PP\bar{A}A\}$.

Interpretation:

- $\bar{P}P\bar{A}\bar{A}$ — effect of **participation only**;
- $\bar{P}P\bar{A}A$ — effect of **participation + benefit**;
- $PP\bar{A}A$ — effect of **benefit only**.

Aggregated effect across relevant strata:

$$\mathrm{ATE}_t^{(\ell)} = \sum_{g \in \mathcal{G}^*} \Pr(G_i = g \mid i \in U_{s0}) \cdot \mathrm{PCE}_{g,t}^{(\ell)}$$

where $\mathcal{G}^* = \{\bar{P}P\bar{A}\bar{A}, \bar{P}P\bar{A}A, PP\bar{A}A\}$.

Rationale: this weighted average represents the causal effect for the **compliers subpopulation**, i.e., firms whose participation or benefit status is affected by eligibility.

Heterogeneity via Semantic Classification

Purpose: Capture heterogeneous causal effects by innovation strategy.

Method: Semantic classification of project texts (2014 call), validated through manual review.

Two orientations:

- **Efficiency-oriented digitalization:** internal optimization, automation, quality control.
- **Expansion-oriented digitalization:** new products, scaling, or market expansion.

Scope of analysis:

- Conducted for firms belonging to the three **principal strata** where eligibility affects participation or benefit: $\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, and $PP\bar{A}A$.
- Project information is observable only when firms participate. For $\bar{P}P\bar{A}\bar{A}$ and $\bar{P}P\bar{A}A$, participation occurs only if firms are **eligible (dynamic)**; thus, heterogeneity inference concerns **participating dynamic firms**.
- In $PP\bar{A}A$ (always participants, beneficiaries only if dynamic), projects exist under both eligibility states; in principle, heterogeneity could extend to non-dynamic firms, but small sample sizes prevent reliable analysis.

Observed Groups (Z, P, A) and Mixtures of Principal Strata

Monotonicity-consistent strata (6):
Relevant (effect estimated): $\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, $PP\bar{A}A$
With exclusion restriction (effect constrained to 0): $\bar{P}\bar{P}\bar{A}\bar{A}^{\dagger}$, $PP\bar{A}\bar{A}^{\dagger}$, $PPAA^{\dagger}$

Z	P	A	**Observed group**	**Principal strata in the mixture**
0	0	0	Non-dynamic, non-participant	$\bar{P}\bar{P}\bar{A}\bar{A}^{\dagger}$, $\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$
0	1	0	Non-dynamic, participates w/out funding	$PP\bar{A}\bar{A}^{\dagger}$, $PP\bar{A}A$
0	1	1	Non-dynamic, participates and funded	$PPAA^{\dagger}$ (*pure*)
1	0	0	Dynamic, non-participant	$\bar{P}\bar{P}\bar{A}\bar{A}^{\dagger}$ (*pure*)
1	1	0	Dynamic, participates w/out funding	$\bar{P}P\bar{A}\bar{A}$, $PP\bar{A}\bar{A}^{\dagger}$
1	1	1	Dynamic, participates and funded	$\bar{P}P\bar{A}A$, $PP\bar{A}A$, $PPAA^{\dagger}$

Note. Strata marked with † are constrained to have zero treatment effect, but they are retained in the estimation to inform the joint posterior and improve the imputation of unobserved potential outcomes **in the relevant strata** ($\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, $PP\bar{A}A$). Cells $(0, 1, 1)$ and $(1, 0, 0)$ are **pure**; all others combine multiple latent strata.

Estimation and Inference

From mixtures to joint estimation: Observed groups (Z, P, A) are mixtures of latent principal strata. The Bayesian framework jointly estimates stratum membership and outcomes.

Bayesian hierarchical setup

- **Strata membership model:** Bivariate probit system for potential participation and benefit equations, conditional on covariates $\mathbf{W}_i$ and forcing variable ΔS_i:

$$\Pr\big(P_i(0), P_i(1), A_i(0), A_i(1) \mid \mathbf{W}_i, \Delta S_i\big)$$

- **Outcome model:** Dynamic panel with firm random effects and stratum-specific parameters, including an explicit model for the **initial condition** (2014 baseline):

$$\left(Y_{it}^{(\ell)}(z) \mid G_i = g\right) = \tau_{g,t} z + \rho_g Y_{i,t-1}^{(\ell)}(z) + \delta_W W_i + u_i + \xi_{it}, \quad Y_{i,2014}^{(\ell)}(z) \sim f(\mathbf{W}_i)$$

- **Inference:** Joint posterior sampling via MCMC (Gibbs + data augmentation). Posterior summaries provide $\mathrm{PCE}_{g,t}^{(\ell)}$, aggregated $\mathrm{ATE}_t^{(\ell)}$, and $\Pr(\mathrm{effect} > 0)$ for each outcome.

Key message: The model integrates uncertainty in both strata membership and outcomes, yielding probabilistic causal effects under weak identification.

Posterior Size and Aggregation of Relevant Strata

Posterior distribution of strata membership

- The union of the **active strata** ($\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, $PP\bar{A}A$) covers about **40–70%** of the local population U_{s0}, depending on the outcome.
- Within this active share, the *Activated* stratum ($\bar{P}P\bar{A}A$) absorbs most of the posterior probability mass.
- The *Stimulated* (participation only) and *Reinforced* (funding only) strata are nearly empty, indicating that eligibility mostly works through joint activation.

Aggregation rationale

- Because two strata are small, inference focuses on the **joint posterior** over all active strata.
- The aggregated effect merges all causal pathways — participation only, funding only, and both — but it is **largely driven by the dominant one**, where eligibility changes both behavior and outcome.
- In this sense, the aggregate provides a **synthetic yet causally interpretable measure** of how the policy mechanism operates in practice.

Interpretation of the Results on Membership Size

Substantive interpretation

- The policy activates about half of the local population — firms for which eligibility truly changes behavior or outcomes.
- Most of this action occurs in the *Activated* stratum, where eligibility triggers both participation and funding.
- The mechanism is **focused but not systemic**: it works effectively on a well-defined active segment, while leaving part of the system inert.

Policy insight

- The eligibility rule **favored dynamic firms**, assuming that turnover growth signaled readiness to innovate.
- This assumption is **not verifiable**: turnover may reflect short-term market conditions rather than project quality.
- The rule was able to identify a set of reactive policy "clients", but this does not mean it was an appropriate policy criterion.
- Its removal today may broaden access and reshape the activation margin — likely adding some inert firms, but also others that could react if given the chance.

Outcomes of Interest

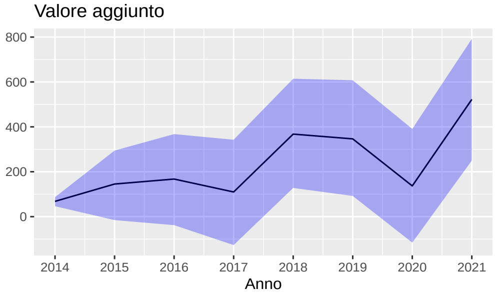
Value added and Workforce Composition

- **Value added** - Productivity inferred jointly from effects on value added and total employees.
- **Employment and workforce composition:**
  - Total employment (headcount).
  - High-skill positions (managers, professionals, technicians).
  - Skilled workers (craftsmen, specialized operators).
  - Plant/machine/vehicle operators.
  - Low-skill positions (elementary occupations).

Value Added: Posterior Effects (Relevant Strata)

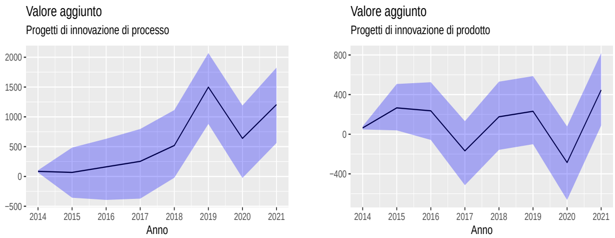
- Posterior mean effects are generally positive over time.
- Pr(effect > 0) often exceeds 0.80, though some 95% credible intervals include zero.

Figure: Causal effect on value added (union of $\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, $PP\bar{A}A$)



Value Added: Efficiency vs. Expansion - seeking projects (Dynamic Firms)

Figure: Causal effects on value added by project type (relevant strata, dynamic firms)

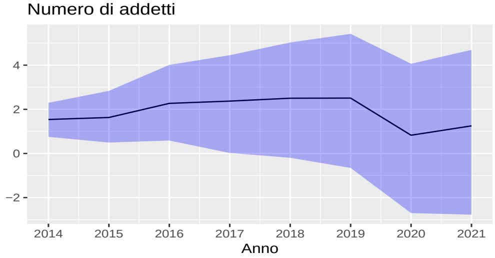


- Effects not statistically significant for *product/expansion* projects.
- More positive (though not always significant) patterns for *process/efficiency-seeking projects*.

Employment: Posterior Effects (Relevant Strata)

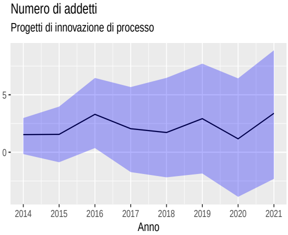
- Small yet generally positive effects on total employment in the medium run.
- Posteriors concentrate on positive values in later years, with wide 95% credible intervals in crisis years.

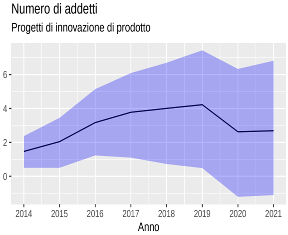
Figure: Causal effect on total employees (union of $\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, $PP\bar{A}A$)



Employment: Efficiency vs. Expansion - seeking projects (Dynamic Firms)

Figure: Causal effects on total employees by project type (relevant strata, dynamic firms)

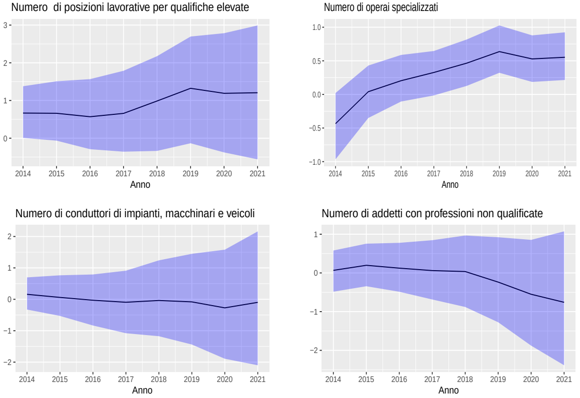




- Non-significant effects for *efficiency-seeking* projects.
- Positive and more often significant effects for *expansion-seeking* projects.

Workforce Composition (Relevant Strata)

Figure: Causal effects by qualification group (union of $\bar{P}P\bar{A}\bar{A}$, $\bar{P}P\bar{A}A$, $PP\bar{A}A$)



- No clear effects on low-skill and operators.
- Predominantly positive effects for specialized and high-skill positions.

Workforce Composition: Automation vs Digitalization (Dynamic Firms)

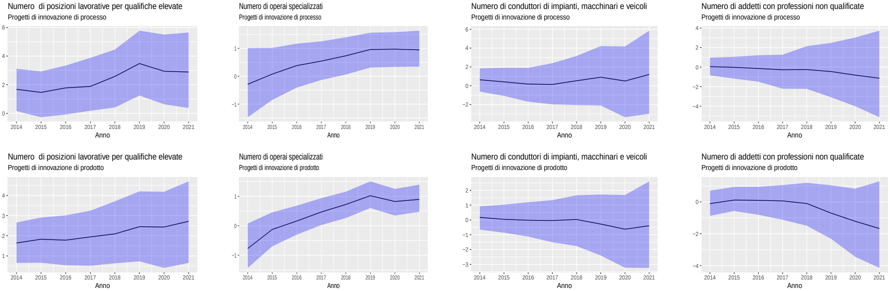


Figure: Causal effects by qualification group and project type (relevant strata, dynamic firms)

- Positive impacts on specialized and high-skill positions in both types.
- No signs of displacement for low-skill .

Conclusions

Efficiency-seeking (process) projects

- Positive and credible effects on **productivity (value added per worker)**.
- Weak or null effects on employment, but upgrading in skill composition.
- Firms "do better" rather than "do more".

Expansion-seeking (product) projects

- Positive effects on **employment**, especially for qualified and specialized workers.
- Limited or no effect on productivity.
- Firms "do more" rather than "do better".

Overall: a policy trade-off between internal efficiency and outward expansion.

Policy takeups

"We are mostly funding R&D staff to speculate in new products, yet productivity barely moves."

The policy logic

- **Product innovation** often creates private returns that firms can widely capture: new products can be priced, branded, or patented.
- **Process innovation**, instead, strengthens efficiency and capabilities through workers and suppliers, thus modernizing the economy
- These **learning and efficiency spillovers** make process innovation socially valuable.
- Process improvements also serve as an **enabling infrastructure**: they make future product innovations easier to produce, adopt, and scale.

Policy direction:

- Rebalance support toward **efficiency-seeking projects**, where productivity and learning externalities are strongest.
- Treat process innovation as a **collective investment in competitiveness**, not merely as firm-level cost saving.

Thank you!

Questions or comments?

Appendix: Logical Space of Principal Strata

$P(0)$	$P(1)$	$A(0)$	$A(1)$	**Status**	**Reason / Label**
0	0	0	0	✓	Unresponsive (inert)
0	0	0	1	×	*Impossible – benefit without participation*
0	0	1	0	×	*Impossible – benefit without participation*
0	0	1	1	×	*Impossible – benefit without participation*
0	1	0	0	✓	Stimulated (active: participation only)
0	1	0	1	✓	Activated (active: participation + benefit)
0	1	1	0	×	*Impossible – benefit without participation under* $Z = 0$
0	1	1	1	×	*Impossible – benefit without participation under* $Z = 0$
1	0	0	0	×	*Excluded – violates monotonicity* $(P(1) < P(0))$
1	0	0	1	×	*Excluded – violates monotonicity* $(P(1) < P(0))$
1	1	0	0	✓	Aspirational (inert: always participate, never funded)
1	1	0	1	✓	Reinforced (active: benefit only)
1	1	1	0	×	*Excluded – violates monotonicity* $(A(1) < A(0))$
1	1	1	1	✓	Self-sufficient (inert: always participate, always funded)
1	0	1	1	×	*Excluded – violates both policy & monotonicity*

Note. Of 16 theoretical combinations, 7 are impossible by policy design ($A(z) > P(z)$), and 3 more are excluded by monotonicity ($P(1) < P(0)$ or $A(1) < A(0)$), leaving 6 consistent strata: 3 active (Stimulated, Activated, Reinforced) and 3 inert (Unresponsive, Aspirational, Self-sufficient).