Three Headlines from One Equation: A Replication and Sensitivity Analysis of "The AI Layoff Trap" (Falk and Tsoukalas, 2026)

Jeremy McEntire*

April 2026

Draft for comment

Abstract

Falk and Tsoukalas (2026) present a task-based automation model in which a demand externality traps firms in an automation arms race, concluding that only a Pigouvian automation tax can correct the distortion. We replicate their model, verify all ten propositions, and confirm the internal logic. We then show that the model's *conclusion* is not a theorem about the world but a conditional statement whose sign and magnitude are determined entirely by parameter choices the authors present as innocuous. Under their baseline ($\eta = 0.30$, zero owner spending, homogeneous goods, no reinstatement feedback), the model produces catastrophe. Under equally defensible parameters reflecting standard features of real economies ($\eta = 0.50$, modest owner spending, product differentiation, endogenous reinstatement), the same model produces stability or even under-automation. Under optimistic but historically grounded parameters ($\eta > 1$, strong reinstatement), it prescribes an automation *subsidy*. Three parameterizations, three headlines, one equation. The math does not adjudicate between them; the inputs do. We further document that the model's baseline fixed-N, fixed-task structural closure—partially relaxed in FT's Section 5.2 endogenous-entry extension, where the trap result persists—does not capture the empirically dominant margins of adjustment: U.S. business applications run 60% above 2019, AI startups captured 52% of global VC

*Independent researcher. Contact: `jandrewmcentire@gmail.com`. Code and replication materials available at the companion repository.

in 2025, and 60% of current U.S. employment is in occupations that did not exist in 1940. FT's endogenous entry allows more firms of the same type; the data show new firm *types* creating new occupational categories. The demand externality mechanism is a genuine theoretical contribution. The policy prescription is a parameter choice.

Keywords: artificial intelligence, automation, labor displacement, robustness analysis, demand externality

JEL Codes: J23, L13, O33, H23

Contents

1 Introduction **5**

2 Model Summary **6**

3 Verification **7**

4 Perturbation Analysis **7**

4.1 Six Channels, One Direction 7

4.2 Combined Perturbations: The Core Result 8

4.3 Monte Carlo Sensitivity 9

5 Real-World Calibration **9**

5.1 Parameter Estimates 10

5.2 Three Diagnoses from One Model 10

5.3 Sector-Level Heterogeneity 10

6 Beyond Endogenous Entry: Empirical Evidence **11**

6.1 Business Formation Has Broken with the Pre-Pandemic Regime 11

6.2 AI Startup Formation Dominates Venture Capital 12

6.3 1.3 Million New AI Jobs in Two Years 12

6.4 The Historical Base Rate 12

7 The Net Employment Evidence **13**

7.1 Aggregate Forecasts 13

7.2 The Augmentation Channel 13

7.3 The Wage Premium 14

7.4 The Wealth Effect and Indirect Job Creation 14

8 Where the Policy Conclusion Breaks Down **14**

9 Discussion **15**

9.1 What FT Get Right 15

9.2 What FT Get Wrong 15

9.3 What We Might Get Wrong 16

9.4 The Entry-Level Paradox 16

9.5 The Missing Variable: Demographics 16

9.6 The Broader Lesson 17

10 Conclusion **17**

References **19**

A Replication Code **21**

1 Introduction

Falk and Tsoukalas (2026, hereafter FT) present a task-based model of automation in which symmetric firms, competing in a single sector, each choose what fraction of their workforce to replace with AI.[1] Their central result is sharp: a demand externality traps rational firms in an automation arms race that harms both workers and owners, and *only* a Pigouvian automation tax can correct it. Universal basic income, capital taxation, worker equity, upskilling, and Coasian bargaining all fail. The paper has received substantial attention, with commentary frequently presenting its conclusions as "mathematically proven."

We take the mathematics seriously and find it correct. What we dispute is the rhetorical move from "this equation yields this result under these assumptions" to "this is what will happen." A model with a formula is not a proof about the world. It is a proof about what happens when you assume the world looks like your model. Newton's $F = Gm_1m_2/r^2$ gives you exactly the gravitational force—once you have decided what m_1, m_2, and r are. The equation is not in dispute. The question is whether your inputs describe anything real.

We demonstrate this concretely by producing three headlines from one equation:

- **Headline A (FT's baseline):** $\eta = 0.30$, zero owner spending, homogeneous goods, no reinstatement. Result: catastrophic over-automation trap. Policy: automation tax.
- **Headline B (moderate realism):** $\eta = 0.50$, modest owner spending, some differentiation, moderate reinstatement. Result: negligible wedge. Policy: monitor but do not intervene.
- **Headline C (historical optimism):** $\eta > 1$, strong reinstatement (the historical base rate). Result: firms automate *too slowly*. Policy: automation *subsidy*.

All three are produced by the same model with the same algebra. The math does not choose between them. Each parameterization is defensible; none is "the answer." The contribution of this paper is to make the parameter dependence explicit and to confront each parameterization with the available empirical evidence.

Specifically, we:

(i) **Replicate and verify** all ten propositions and four corollaries numerically, confirming the model's internal consistency (Section 3);

[1] "The AI Layoff Trap," arXiv:2603.20617v1, posted March 21, 2026. Working paper; not peer-reviewed at time of writing.

(ii) **Perturb** the baseline assumptions along six empirically motivated dimensions, showing that the over-automation wedge is highly sensitive to each and can be eliminated or inverted by combining modest relaxations (Section 4);

(iii) **Calibrate** to observable parameters, showing that different defensible choices produce different diagnoses—none of which should be taken as definitive (Section 5);

(iv) **Document** that the model's structural closure—fixed tasks and within-sector entry, even in the endogenous-entry extension of Section 5.2—does not capture the dominant empirical margins of adjustment (Section 6).

A note on symmetry: the optimistic "Net-Positive Paradigm" narrative—citing WEF's 78 million net jobs, PwC's wage premia, and Goldman's wealth effect—is as much a parameter choice as the "Layoff Trap" narrative. We present both the optimistic and pessimistic evidence not to adjudicate between them but to bound the parameter space from both sides. Our conclusion is not that the economy is fine. Nor is it that the economy is doomed. It is that the model cannot distinguish between these states without empirical calibration that the paper does not provide, and that the structural margins the model assumes away—new firms, new tasks, new occupations—are precisely where the action is in the current AI transition.

2 Model Summary

We briefly restate FT's model to fix notation. $N \geq 2$ symmetric firms each have L task-positions. Firm i automates fraction $\alpha_i \in [0,1]$, saving $s = w - c$ per task but incurring convex integration cost $\frac{k}{2} L \alpha_i^2$. Workers spend fraction λ of income in the sector; owners spend zero. A fraction η of displaced wage income is replaced. The effective demand loss per automated task is $\ell = \lambda(1-\eta)w$.

Revenue is split equally: $\text{Rev}_i = D/N$ where $D = A + \lambda w L N[1-(1-\eta)\bar{\alpha}]$. Firm i's profit is:

$$\pi_i = \Pi_0 + L\left[\alpha_i\left(s - \frac{\ell}{N}\right) - \frac{k}{2}\alpha_i^2 - \frac{\ell}{N}\sum_{j\neq i}\alpha_j\right]. \tag{1}$$

The Nash equilibrium rate is $\alpha^{NE} = \min\left(\frac{s-\ell/N}{k}, 1\right)$ when $N > N^* = \ell/s$, and the cooperative optimum is $\alpha^{CO} = \min\left(\max\left(0, \frac{s-\ell}{k}\right), 1\right)$. The over-automation wedge is:

$$\alpha^{NE} - \alpha^{CO} = \frac{\ell(1-1/N)}{k} \tag{2}$$

when both are interior. This wedge is the paper's central object.

3 Verification

We implemented all equations, propositions, and extensions from FT in Python. At the paper's illustrative parameters ($c/w = 0.30$, $\lambda = 0.5$, $\eta = 0.30$, $N = 7$, $k = 1$, $A = 10$), our numerical results match the analytical formulas to machine precision ($< 10^{-6}$) across all ten propositions and four corollaries. Specifically:

- **Proposition 1:** $\alpha^{NE} = 0.650$, $\alpha^{CO} = 0.350$, wedge $= 0.300$. Dominant-strategy property verified by grid search over random opponent profiles.
- **Corollary 1:** At $k = 0$, full automation is strictly dominant when $N > N^*$. Deadweight loss $= NL(\ell - s)$ confirmed.
- **Proposition 2:** Surplus loss formula verified for $\mu \in \{0, 0.2\}$. Pareto dominance confirmed: both $\mathcal{K}$ and $\mathcal{W}$ are lower at α^{NE} than at α^{CO}.
- **Propositions 3–10:** Worker equity, coalitions, Pigouvian tax, AI productivity, endogenous entry, endogenous wages, and capital recycling all verified numerically.

The math is correct. Everything that follows concerns the assumptions, not the algebra.

4 Perturbation Analysis

The over-automation wedge $\ell(1 - 1/N)/k$ depends on three quantities: the demand loss ℓ, the number of firms N, and integration friction k. We show that ℓ is highly sensitive to assumptions that FT treat as innocuous baseline choices.

4.1 Six Channels, One Direction

We perturb six baseline assumptions independently. Each operates on the effective demand loss ℓ through a different economic mechanism; all shrink the wedge. We describe each channel's logic, then present the quantitative results together in Table 1.

Non-zero owner consumption ($\hat{\eta} > 0$). FT normalize owner sectoral spending to zero. But owners are also consumers. When owners spend fraction $\hat{\eta}$ of profits in the sector, each automated task's demand loss shrinks from ℓ to $\ell_{\hat{\eta}} = \ell - \hat{\eta}s$. The Nash rate shifts to $(s - \ell/\hat{N})/k$ where $\hat{N} = N(1 - \hat{\eta}) + \hat{\eta}$, making each firm behave as though it faced fewer competitors.

Endogenous income replacement ($\gamma > 0$). FT fix $\eta = 0.30$. But the Acemoglu–Restrepo reinstatement effect implies η responds endogenously to automation: new complementary tasks and roles emerge as firms automate. We model $\eta(\bar{\alpha}) = \eta_0 + \gamma\bar{\alpha}$, capped at 1. At $\gamma = 1.0$—matching the long-run historical reinstatement rate—the wedge vanishes entirely. FT's result requires $\gamma = 0$: zero feedback from automation into income replacement.

Product differentiation ($\delta > 0$). FT assume $\text{Rev}_i = D/N$: perfectly homogeneous goods. With differentiation, a firm that automates captures a larger revenue share, internalizing more of its own demand loss. We parameterize internalization as $1/N + \delta(1 - 1/N)$, where $\delta = 0$ is the paper's baseline. At $\delta = 0.30$ (modest brand loyalty), the wedge drops 30%.

Consumption smoothing ($\sigma > 0$). FT's static model assumes displaced workers immediately and fully cut spending. Workers with savings, credit, or severance smooth consumption over time. Modeling the effective demand loss as $\ell(1-\sigma)$, even $\sigma = 0.30$ (plausible over a one-year horizon) cuts the wedge by 30%.

Heterogeneous firms. FT's symmetry assumption forces the Prisoner's Dilemma structure. With heterogeneous AI costs, markets self-segment: AI-native firms automate heavily while human-service-intensive firms rationally choose the human niche. This "division of labor" response is ruled out by construction in the symmetric model.

Non-linear integration costs. FT use quadratic friction $\frac{k}{2}\alpha^2$. Under an S-curve—cheap early automation, steep middle, plateauing at full—the equilibrium structure changes qualitatively. Under marginal costs $c'(\alpha) = k \cdot 4\alpha(1 - \alpha)$, both NE and CO shift to full automation and the wedge collapses.

Table 1: Wedge sensitivity to individual perturbations (FT illustrative parameters; only the perturbed channel varies, others at baseline)

Channel	Perturbation	Wedge	Reduction
Baseline (FT)	—	0.300	—
Owner MPC	$\hat{\eta} = 0.40$	0.274	8.7%
Reinstatement	$\gamma = 0.50$	0.207	31%
Differentiation	$\delta = 0.30$	0.210	30%
Smoothing	$\sigma = 0.30$	0.210	30%
S-curve costs	see text	0.000	100%

The individual channels are modest. None alone is devastating. Their force is cumulative.

4.2 Combined Perturbations: The Core Result

Each perturbation alone is modest. Combined, they are devastating.

Table 2: Combined perturbations at FT's illustrative parameters

Scenario	$\hat{\eta}$ / γ / σ / δ	α^{NE}	α^{CO}	Wedge	vs. Baseline
Paper baseline	0 / 0 / 0 / 0	0.650	0.350	0.300	100%
Conservative	0.15 / 0.20 / 0.20 / 0.15	0.667	0.571	0.096	32%
Moderate	0.30 / 0.40 / 0.30 / 0.25	0.724	0.773	−0.050	inverted
Strong	0.40 / 0.60 / 0.40 / 0.35	0.794	0.939	−0.145	inverted

Result 1 (Fragility)**.** *At the "conservative realism" parameterization—each perturbation set to 15–20% of its feasible range—the over-automation wedge drops from 0.300 to 0.096, a 68% reduction. At "moderate realism," the wedge inverts: firms under-automate relative to the cooperative optimum.*

None of these assumptions are controversial. Non-zero owner spending, endogenous reinstatement, product differentiation, and consumption smoothing are standard in the literatures FT cite.

4.3 Monte Carlo Sensitivity

To ensure these results are not an artifact of the illustrative parameter vector, we draw 10,000 random parameter combinations uniformly over the feasible space ($N \in [2, 50]$, $c \in [0.05, 0.95]$, $\lambda \in [0.1, 1.0]$, $\eta \in [0, 0.8]$, $k \in [0.1, 3.0]$).

Table 3: Fraction of parameter space with wedge > threshold, $n = 10{,}000$

Threshold	Baseline	With perturbations	Change
0.05	78.5%	39.6%	−49.6%
0.10	57.3%	22.5%	−60.7%
0.20	30.9%	8.6%	−72.2%
0.30	17.9%	4.2%	−76.7%

The mean wedge drops from 0.189 to 0.066. The fraction of parameter space exhibiting a "large" wedge (> 0.20) falls from 31% to under 9%. The model's catastrophic result occupies a small, fragile corner of its own parameter space.

5 Real-World Calibration

We calibrate the model to observable U.S. economic parameters, drawing on BLS data, sector-level concentration ratios, and the labor economics literature.

5.1 Parameter Estimates

Table 4: Best-estimate current parameters

Parameter	Symbol	Estimate	Source/Rationale
AI cost ratio	c/w	0.35	Compute costs + training; not free
Worker MPC	λ	0.55	US savings rate 3–5%; multi-sector dilution
Income replacement	η	0.50	UI + retraining + new job creation
Firms	N	15	Varies by sector; median estimate
Integration friction	k	1.2	Trust, regulation, organizational inertia
Owner MPC	$\hat{\eta}$	0.25	Owners consume across sectors
Reinstatement rate	γ	0.30	Historical new-task creation rate
Consumption smoothing	σ	0.25	Savings, credit, severance
Product differentiation	δ	0.20	Brand loyalty, switching costs

5.2 Three Diagnoses from One Model

The point is not which diagnosis is "correct" but that the model produces any of them depending on choices the modeler must make before running it.

Diagnosis A (FT's parameters, no perturbations): $\alpha^{NE} = 0.526$, $\alpha^{CO} = 0.313$, wedge $= 0.214$. *Headline: Concerning over-automation; tax warranted.*

Diagnosis B (Table 4 parameters, with perturbations): $\alpha^{NE} = 0.547$, $\alpha^{CO} = 0.563$, wedge $= -0.017$. *Headline: Negligible wedge; no intervention needed.*

Diagnosis C (same as B but $\eta = 0.70$, $\gamma = 0.60$): wedge ≈ -0.10. *Headline: Firms under-automate; subsidy warranted.*

Result 2 (Parameter Dependence)**.** *The model's policy prescription—tax, do nothing, or subsidize—is determined entirely by the modeler's choice of η, γ, $\hat{\eta}$, δ, and σ. None of these quantities is directly observable. Each diagnosis above uses defensible values; none is "the answer." The honest conclusion is that the sign of the optimal intervention is an empirical question the model cannot resolve without external calibration data that neither this paper nor FT provides with sufficient precision.*

5.3 Sector-Level Heterogeneity

The model's predictions vary dramatically across sectors:

The paper's trap is most plausible in fragmented, commoditized sectors with low-skill workers (retail)—exactly where automation has *always* been most disruptive. It is least

Table 5: Sector calibrations: baseline vs. realistic wedge

Sector	N	Baseline wedge	Realistic wedge	Status
Customer Support	20	0.200	0.000	Stable
Software Engineering	8	0.102	−0.079	Stable
Financial Back-Office	12	0.238	−0.070	Stable
Retail	30	0.725	0.563	Vulnerable
Creative/Media	50	0.000	0.000	Boundary

plausible in concentrated, differentiated sectors with skilled, mobile workers (software, finance).

6 Beyond Endogenous Entry: Empirical Evidence

FT's baseline model fixes the number of firms N, the task set L, and output $Y_i = L$. Their Section 5.2 partially relaxes this by modeling endogenous entry: firms enter until profits net of a fixed cost κ reach zero, endogenizing N. This is a meaningful extension, and we verified it numerically (Section 3). Crucially, however, the over-automation trap *persists* under endogenous entry in FT's framework—more firms enter, but each still faces the same demand externality, and the wedge remains positive at the free-entry equilibrium.

Moreover, FT's endogenous entry adds more firms of the *same type* to the *same sector*. The empirical phenomenon we document below is qualitatively different: new firm types creating entirely new occupational categories and sectors. The task set L remains fixed throughout all of FT's extensions, which eliminates the channel—new task creation—that the empirical literature identifies as the dominant margin of labor-market adjustment to technology shocks.

6.1 Business Formation Has Broken with the Pre-Pandemic Regime

The Census Bureau's Business Formation Statistics show a structural break persisting for more than five years. Total business applications rose from approximately 3.5 million in 2019 to 5.67 million in 2025—a sustained 60% increase. High-propensity business applications (those most likely to become employer firms) averaged 142,000 per month in 2024, more than 40% above the pre-pandemic baseline.

Decker and Haltiwanger (2024), the economists most identified with measuring U.S. business dynamism, find that high-tech industries contributed disproportionately to the post-2019 entry surge, specifically calling out "research and development services such as artificial

intelligence businesses." FT's endogenous-entry extension allows N to adjust within a sector, but the new entrants are identical to incumbents. It cannot represent the phenomenon in which AI cost reductions capitalize into new *types* of firms that hire workers for new tasks.

6.2 AI Startup Formation Dominates Venture Capital

Y Combinator's AI share rose from 15% in Summer 2022 to 88% in Summer 2025. Stanford's AI Index reports 1,953 newly funded AI companies in the U.S. in 2025 alone. Private AI investment reached $285.9 billion; total corporate AI investment surpassed $500 billion—double 2024. AI accounted for roughly 52% of global VC (PitchBook, 2025). CB Insights counts 75 new AI unicorns in 2025, representing 61% of all new unicorn births.

Stripe Atlas data capture the small-founder tail: the share of Atlas startups building AI rose from 15% (January 2023) to 42% (2025). Solo founders now represent 35% of 2024 incorporations. AI coding tools are the enabler: Cursor reached $2 billion ARR in 14 months at a $29.3 billion valuation; by early 2026, 51% of code committed to GitHub was AI-generated or AI-assisted.

6.3 1.3 Million New AI Jobs in Two Years

LinkedIn's January 2026 Economic Graph report documents 1.3 million new AI-related jobs globally in 2023–2025, with AI-related hiring nearly doubling since 2023. "Head of AI" postings tripled over five years; Chief AI Officer adoption across the Fortune 500 climbed from 11% (2023) to 26% (2025). Indeed Hiring Lab measures AI postings at 4.2% of all U.S. job listings in December 2025—an all-time high.

Stanford AI Index/Lightcast data show U.S. postings mentioning prompt engineering rising from 1,400 (2023) to 6,300 (2024), LLM mentions from 5,000 to 20,000, generative-AI skill postings from 16,000 to 66,000, and agentic-AI mentions up 280% year-over-year. Scale AI alone employs roughly 5,800 FTEs plus 240,000+ contractor annotators—an entire occupational ecosystem that was statistically negligible before 2022.

6.4 The Historical Base Rate

Autor, Chin, Salomons, and Seegmiller (2024) establish that roughly 60% of 2018 U.S. employment was in occupations that did not exist in 1940. Among professionals, the figure is 74%. FT cite this paper but use it only to calibrate $\eta < 1$, discarding its central finding: new-work creation is endogenous to the technology shock.

Every prior general-purpose technology wave exhibited the same pattern. Spreadsheets destroyed roughly 400,000 accounting-clerk jobs since 1980 but added approximately 600,000 accountant positions plus financial analysts and FP&A roles. Bessen's 1950–2010 study found that only one of 271 detailed occupations (elevator operators) was eliminated largely by automation. ATMs reduced per-branch tellers from 20 to 13, but urban bank branches rose 43% and total teller employment increased. The smartphone wave generated approximately 2.4 million U.S. iOS app-economy jobs by 2023—an occupation set that was literally zero in 2007.

FT's endogenous-entry extension allows N to adjust, but the fixed task set L and within-sector structure remain. The assumption that does the most work is not fixed N per se—it is the absence of new task creation and cross-sector reallocation, channels that the data show operating at historically elevated rates.

7 The Net Employment Evidence

The macroeconomic consensus across major global institutions projects net-positive AI employment effects, contradicting the paper's framing of inevitable demand collapse.

7.1 Aggregate Forecasts

Table 6: Institutional employment forecasts

Source	Projection	Net Impact	Horizon
WEF (2025)	170M created, 92M displaced	+78M jobs	2030
U.S. BLS	+5.2M new jobs	+3.1% growth	2024–34
BCG (2026)	50–55% reshaped; 10–15% eliminated	Net augmentation	2026–31
Goldman Sachs	300M exposed	Productivity + infrastructure offsets	10-year

The WEF *Future of Jobs Report 2025*, surveying over 1,000 employers representing 14 million workers across 55 economies, projects a net gain of 78 million jobs by 2030. The U.S. BLS projects 5.2 million new jobs through 2034, contradicting narratives of mass unemployment. Critically, the deceleration in growth rates is attributed to demographic constraints—an aging workforce—not technological displacement.

7.2 The Augmentation Channel

The ILO's 2025 analysis of 52,558 data points across 30,000 tasks finds that while one in four workers globally has some GenAI exposure, the vast majority of these jobs will be

transformed rather than eliminated. Employer demand for augmentation-prone roles (where AI enhances rather than replaces human output) grew 20% (Harvard Business School, 2025), even as postings for pure-automation-exposed roles declined 13%.

7.3 The Wage Premium

PwC's 2025 AI Jobs Barometer, analyzing nearly one billion job advertisements, finds that:

- Industries with highest AI exposure see revenue per employee growing **3**× faster than less exposed industries;
- Wages in AI-exposed industries are rising **twice as fast** (16.7% vs. 7.9%);
- Workers with AI skills command a **56% wage premium** over peers in the same occupations without those skills, up from 25% one year prior.

If AI were a blunt labor-substitution tool, economic theory predicts wage stagnation in exposed sectors. The data show the opposite.

7.4 The Wealth Effect and Indirect Job Creation

AI is driving one of the largest capital expenditure cycles in modern economic history. Barclays estimates that AI infrastructure spending contributed a full 1% of U.S. economic growth in 2025. This physical build-out—data centers, power grids, semiconductor fabs—requires construction workers, electrical engineers, HVAC specialists, and logistics coordinators. Goldman Sachs models the resulting "wealth effect" as a primary pillar of AI job creation: productivity gains generate discretionary income that creates demand for human-centric services (healthcare, personal services, education) that are fundamentally immune to automation.

The U.S. currently employs over 1 million workers in occupations such as pet care, personal training, and specialized tutoring—roles whose existence depends entirely on the discretionary income generated by broader economic prosperity.

8 Where the Policy Conclusion Breaks Down

FT's Pigouvian automation tax ($\tau^* = \ell(1 - 1/N)$) is derived from a model where new-task creation is impossible and new-firm creation, even when endogenized (Section 5.2), adds identical firms to a fixed sectoral structure. In the empirical record, the dominant margins through which technology shocks expand employment are new firm *types* and new

occupational categories—channels the model does not capture. A tax on automation cost reductions risks taxing precisely the mechanism that, in every prior technology wave, has capitalized into new firms and new occupations.

FT concede their result depends entirely on $\eta < 1$, and justify that calibration with Jacobson, LaLonde, and Sullivan's (1993) *within-occupation* earnings-loss estimates. But within-occupation earnings losses cannot speak to *cross-sector* reallocation via new-firm birth—the channel documented by Haltiwanger et al. (2013), who establish that young firms account for essentially all net job creation in the U.S. economy, and by Decker and Haltiwanger (2024), who document this channel operating right now in AI-related industries.

The paper's own framework reveals the fragility of the policy conclusion. By Corollary 2, if $\eta > 1$—displaced workers get *better* jobs—the externality flips and the optimal policy is an automation *subsidy*. At our "moderate realism" calibration, the wedge is already negative. A policymaker taking the model seriously should recognize that the sign of the optimal intervention is an empirical question, not a theoretical one, and the current data lean toward the subsidy side.

9 Discussion

9.1 What FT Get Right

The demand externality mechanism is genuine. In any competitive market where displaced workers' lost spending reduces rival firms' revenue, each firm bears only $1/N$ of the demand destruction it causes. This is a real coordination problem, theoretically isomorphic to pollution externalities, and worth studying. The paper's taxonomy of policy instruments (Table 1) is a useful organizing framework. And the Prisoner's Dilemma structure in the frictionless limit is an elegant illustration.

9.2 What FT Get Wrong

The paper presents one parameterization as the baseline and treats every alternative as a robustness check. This framing creates the impression that the catastrophic result is the default and that deviations from it require justification. The opposite framing—start with observable features of actual economies and ask what must fail for the trap to activate—is equally valid and produces a less alarming picture. But we should be honest: our calibration choices are also choices. The economy does not come labeled with its η.

9.3 What We Might Get Wrong

The optimistic empirical evidence we cite—WEF's 78 million net jobs, PwC's wage premia, the startup boom—is itself subject to selection and framing effects. The WEF survey asks executives who are investing in AI whether AI will create jobs; the PwC barometer measures postings, not hires; startup formation rates measure entry, not survival. The "Net-Positive Paradigm" narrative is as much a parameter choice as the "Layoff Trap" narrative. We present the empirical evidence not as proof of safety but as a constraint on the parameter space—one that makes FT's worst-case baseline harder to defend while not ruling out intermediate scenarios.

9.4 The Entry-Level Paradox

One empirical finding deserves separate treatment because it partially vindicates FT's concerns. Anthropic's 2026 labor market study identifies a localized hiring slowdown for younger workers (aged 22–25) entering highly AI-exposed occupations: a 14% decline in the job-finding rate relative to 2022 baselines, with no comparable decline for workers over 25. This is consistent with firms automating the routine tasks that traditionally served as apprenticeship pathways. The aggregate data remain strong, but the aggregate can mask acute transitional pain at the entry level. A model tuned to entry-level cognitive work in 2025 might well produce a positive wedge—not because the whole economy is trapped, but because one segment of it is experiencing genuine friction. Acknowledging this does not rescue FT's universal policy prescription, but it does identify the sector-specific conditions under which their mechanism has bite.

9.5 The Missing Variable: Demographics

FT's model assumes a labor surplus: every displaced worker adds to a pool that depresses demand. But the advanced economies implementing AI most aggressively face the opposite problem. The OECD *Employment Outlook 2025* documents that old-age dependency ratios have risen from 19% (1980) to 31% (2023) and project 52% by 2060. In this demographic context, automation is not a trap but a mathematical necessity to maintain per-capita output as the workforce shrinks. The model contains no demographic channel; adding one would likely push η above 1 for economies with severe labor shortages, flipping the sign of the externality by the model's own logic (Corollary 2).

9.6 The Broader Lesson

Every mathematical model is a conditional statement: *if* these assumptions hold, *then* this follows. The math maps assumptions to conclusions; it does not validate the assumptions. When commentary describes a model's conclusions as "mathematically proven," it conflates logical validity with empirical truth. FT's theorem is valid. Whether it describes the actual economy is an empirical question that neither their paper nor ours definitively answers. What we establish is that the question is *open*—not settled by the existence of an equation.

10 Conclusion

We replicate Falk and Tsoukalas (2026), verify their mathematics, and confirm the internal logic of all ten propositions. We then demonstrate that the model's *output*—the sign and magnitude of the over-automation wedge, and therefore the policy prescription—is determined by parameter choices that the algebra cannot adjudicate:

1. **Individually fragile:** small perturbations to any one of six baseline assumptions substantially reduce the over-automation wedge.

2. **Jointly decisive:** combining modest perturbations across four channels eliminates or inverts the wedge. The model produces "trapped," "stable," or "under-automating" depending on inputs that are not directly observable.

3. **Structurally constrained:** the fixed-task closure—and the within-sector nature of the endogenous-entry extension (Section 5.2 of FT)—does not capture the margins that the empirical literature identifies as dominant: new firm *types*, new tasks, and new occupations, all of which live data show operating at historically elevated rates.

4. **Historically unusual:** every prior general-purpose technology wave exhibited the reinstatement dynamics the model assumes away.

The demand externality is a real theoretical mechanism, worth studying and worth monitoring. The question of whether it is quantitatively large enough to warrant a Pigouvian automation tax—or small enough to ignore, or negative enough to warrant a subsidy—is an empirical question that no amount of algebra can settle. FT's model provides the conditional logic; it does not provide the calibration. The conclusion lives in the parameters, not the math.

Models are useful precisely because they are incomplete. They isolate one mechanism and show what it implies in the absence of everything else. The error is not in building such a model. The error is in forgetting what was left out.

References

Acemoglu, D. and Restrepo, P. (2018).
"The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review*, 108(6), 1488–1542.

Acemoglu, D. and Restrepo, P. (2019).
"Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives*, 33(2), 3–30.

Autor, D.H. (2015).
"Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives*, 29(3), 3–30.

Autor, D.H., Chin, C., Salomons, A., and Seegmiller, B. (2024).
"New Frontiers: The Origins and Content of New Work, 1940–2018." *Quarterly Journal of Economics*, 139(3), 1399–1465.

BCG (2026).
"AI Will Reshape More Jobs Than It Replaces." Boston Consulting Group.

Bessen, J.E. (2019).
"Automation and Jobs: When Technology Boosts Employment." *Economic Policy*, 34(100), 589–626.

Brynjolfsson, E., Li, D., and Raymond, L. (2025).
"Generative AI at Work." *Quarterly Journal of Economics*, 140(2), 889–942.

Decker, R.A. and Haltiwanger, J.C. (2024).
"Surging Business Formation in the Pandemic Era." *FEDS Notes*, Board of Governors of the Federal Reserve System.

Falk, B.H. and Tsoukalas, G. (2026).
"The AI Layoff Trap." arXiv:2603.20617v1.

Haltiwanger, J.C., Jarmin, R.S., and Miranda, J. (2013).
"Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics*, 95(2), 347–361.

ILO (2025).
"Generative AI and Jobs: A Refined Global Index of Occupational Exposure." International Labour Organization.

IMF (2026).
"Bridging Skill Gaps for the Future: New Jobs Creation in the AI Age." Staff Discussion Note SDN/2026/001.

Jacobson, L.S., LaLonde, R.J., and Sullivan, D.G. (1993).
"Earnings Losses of Displaced Workers." *American Economic Review*, 83(4), 685–709.

PwC (2025).
"AI Jobs Barometer." PricewaterhouseCoopers.

Stanford HAI (2025).
"AI Index Report 2025." Stanford Institute for Human-Centered AI.

WEF (2025).
"Future of Jobs Report 2025." World Economic Forum.

Yale Budget Lab (2026).
"Evaluating the Impact of AI on the Labor Market: January/February CPS Update."

Anthropic (2026).
"Labor Market Impacts of AI: A New Measure and Early Evidence."

OECD (2025).
"Employment Outlook 2025: Can We Get Through the Demographic Crunch?"

A Replication Code

All code for the verification, perturbation analysis, and calibration is available as a Python package. The implementation consists of three files:

- `model.py`: Core equations for all propositions and extensions.
- `verify.py`: Numerical verification of Propositions 1–10.
- `perturb.py`: Perturbation analysis and Monte Carlo sensitivity.
- `calibrate.py`: Real-world calibration and sector analysis.

All results in this paper can be reproduced by running these scripts with Python 3.10+ and NumPy/SciPy.