UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉPARTEMENT DES SCIENCES ADMINISTRATIVES

WORKING PAPER NO. 25

When AI Agents Become Economic Agents

An Agentic General Equilibrium Theory

Simon-Pierre Boucher

Département des sciences administratives
Université du Québec en Outaouais

simon-pierre.boucher@uqo.ca

Gatineau – Pavillon Alexandre-Taché
283, boulevard Alexandre-Taché
Gatineau, Québec, Canada J9A 1L8

October 2026

Version 1.0

Abstract

Households and firms are beginning to delegate economic decisions to autonomous agents built on a few shared foundation models. We study what this does in general equilibrium. A human's decision error is idiosyncratic and averages out; an AI agent's error is common to every principal that uses the same model. Delegation therefore converts noise that cancels into noise that does not: capital accumulation, asset demand and credit load on the shared error with weight D (the delegation share) and $1/M$ (the effective number of independent models), and wages, interest rates and asset prices inherit this "model noise". In a task-based stochastic growth economy with a compute market, the decision problem reduces to a linear-quadratic game with dispersed information in which delegation is the purchase of a correlated signal. We prove that in a price-taking economy with constant returns the private value of delegation equals its social marginal value identically—model noise is real but not an inefficiency—and we obtain closed-form wedges, Pigouvian delegation taxes and the rate at which model diversity substitutes for them when the aggregate action has non-pecuniary effects, when price risk is uninsurable, or when prices enter collateral constraints. Delegating the consumption–savings decision has almost no aggregate bite; delegating the allocation of wealth has a large one. In a heterogeneous-agent economy the rich delegate first, better decisions concentrate wealth even with broad ownership of AI, and the distributional effect of automation is decided by who owns the compute. In an agentic market laboratory with explicit balance sheets, margin rules, defaults, three market designs and live open-weight language-model agents under hard deterministic constraints, order-flow correlation rises with D and falls with M in every design; a shared model makes markets quieter and more correlated at once; and a false shared signal is amplified by collateral in a call market (by a factor of 2.6 at a 20% margin, halved at 50%, eliminated by model diversity) but not in a continuous auction. Language-model agents herd less than Bayesian agents on the same shared signal but load prices on it more, and their proposed actions routinely violate the constraints that deterministic code enforces. Code, data provenance, prompts and audit logs are provided.

Keywords: artificial intelligence, delegation, foundation models, dispersed information, general equilibrium, compute, wealth inequality, market microstructure, LLM agents

JEL Classification: D82, D83, E13, E21, E44, G12, G14, O33

1. Introduction

Economists have studied artificial intelligence mostly as something that is *produced with*: a capital good that performs tasks, a technology that raises productivity, a force that displaces labour. This paper studies AI as something that *decides*. Households and firms are beginning to delegate economic decisions—how to allocate savings, when to trade, what to order, how much to borrow—to autonomous agents built on a handful of foundation models. The question we ask is what changes in general equilibrium when the decisions themselves, and not only the tasks, are made by machines that many principals share.

The mechanism we isolate is simple and, as far as we can tell, has not been written down. Economic decisions are taken on forecasts of payoff-relevant states. A human's forecast error is idiosyncratic: across a continuum of households it washes out by the law of large numbers, so it costs only the person who makes it. A decision delegated to an agent built on foundation model m carries model m's error, which is *common* to every principal that uses the same model. Delegation therefore has two faces. Privately, it raises forecast precision and saves time, at a fee priced in the compute market. In the aggregate, it replaces a noise that averages out with one that does not: capital accumulation, asset demand and credit start to load on the shared error with weight D, the delegation share, and $1/M$, the effective number of independent models. Wages, interest rates and asset prices inherit what we call *model noise*. A market for delegation thus sets the error-covariance structure of the economy.

Two results organise the paper. The first is a benchmark that goes against the most common policy narrative. In a price-taking economy with constant returns and complete markets, the private value of delegating a decision equals its social marginal value *identically*—for every delegation share, every number of models and every degree of strategic interaction (Proposition 3). Model noise is real, but it is not an inefficiency: delegators, knowing that others act on the same signal and that capital accumulation is a game of strategic substitutes, dampen their response exactly as a planner would, and the pecuniary effects of their common error on everyone else are zero-sum by Euler's theorem. We derived this symbolically before running any simulation, and we present it first because a referee should see it first. The second result characterises when delegation *does* become a policy problem: when the aggregate action has non-pecuniary effects, when the price risk created by model noise falls on agents who cannot insure it, or when prices enter collateral constraints (Propositions 4 and 7). In each case we obtain the sign and the size of the wedge between private and social value in closed form, the corresponding Pigouvian delegation tax, and the rate $(1/M)$ at which model diversity substitutes for it. In the benchmark, diversity is irrelevant; under the frictions, it is a substitute for the tax.

The economy is built in three nested models so that each claim is made at the level of abstraction

where it can be proved or tested. Model A is a task-based stochastic growth economy (Acemoglu and Restrepo, 2018) with a compute market in which production automation and inference for delegated decisions compete for the same stock. Its within-period decision problem reduces to a linear-quadratic game with dispersed information in the class of Angeletos and Pavan (2007), where delegation is the purchase of a correlated signal; this is where the propositions live. Two findings come out of mapping the economy to the game. Delegating the consumption–savings decision has almost no aggregate bite (a one-log-point productivity news moves desired savings by less than 0.002), because wealth and substitution effects cancel; delegating the allocation of wealth between risky and safe assets has a large one. Which decision is delegated matters more than how good the model is. With illustrative information parameters, the first household to delegate its allocation decision gains about 3% of consumption per year, the last gains 0.4%, and at full adoption with one model the model-noise component of the wage has a standard deviation of 7.7% a year—larger than what calibrated productivity shocks generate.

Model B is a stationary heterogeneous-agent economy (Aiyagari, 1994) in which the decision channel appears as a decision-quality return term and compute is owned broadly or by a small class. Three facts emerge. The rich delegate first, because the stake is proportional to wealth and part of the fee is not. The information channel alone raises the Gini coefficient from 0.68 to 0.70 and the top-10% wealth share from 46% to 47% while adding only 3% to output: better decisions are a scale-dependent return, and scale-dependent returns concentrate wealth (Gabaix et al., 2016). AI as a production input works the other way under broad ownership—wages rise, returns fall, the Gini falls to 0.58—and the other way again under concentrated ownership (Gini 0.66, top-10% share 47%). The distributional consequence of AI is decided by who owns the algorithmic capital, not by the technology.

Model C replaces the auctioneer with explicit trading, balance sheets, collateralised credit and defaults, and populates the market with four families of agents: a Bayesian mean-variance control, rules, a deterministic surrogate behind the LLM interface, and live open-weight language models (Qwen2.5-7B, Mistral-7B, Llama-3.2-3B) whose actions are schema-validated and clipped by deterministic code before they touch any ledger. In a factorial design over market designs (Walrasian call, continuous double auction, bilateral bargaining), delegation share and number of models with common random numbers, the order-flow correlation conditional on public information rises from 0.11 with private signals to 0.97 with one shared model and falls to 0.43 with eight, in every design. The competing explanation—homogenisation removes disagreement and therefore trade—is also present: volume collapses toward the no-trade outcome and unconditional volatility does not rise. Fragility appears conditional on a shock. In the call market, a false shared signal that moves the price by 3.7% when agents are unlevered moves it by 9.8% when they start at 95% of their margin capacity with a 20% margin, an amplification of 2.6; a 50% margin halves it; four

models eliminate it. In the continuous auction the same shock produces a 10% trough and almost no cascade, because sequential execution lets the price recover before the next margin check. Whether model noise is amplified by collateral depends on how execution turns correlated orders into a marked price. Language-model agents given the same shared signal herd much less than Bayesian agents (their outputs add idiosyncratic noise to a common signal), yet their trades load prices on the shared error more, they keep trading when signals are identical, and half to three-quarters of their proposed actions hit the cash, holdings or margin limits.

Relation to the literature. Common-noise fluctuations, the social value of public information and the efficiency of the use of information in neoclassical economies are classical (Lorenzoni, 2009; Angeletos and La'O, 2013; Morris and Shin, 2002; Angeletos and Pavan, 2007); the endogenous purchase of the same information has been studied (Veldkamp, 2006; Hellwig and Veldkamp, 2009; Myatt and Wallace, 2012); the portfolio externality of crowded strategies is known (Stein, 2009; Wagner, 2011); and the closest recent work shows that AI adoption with correlated signals degrades price efficiency in a Kyle market (Meng and Chen, 2026) or states our premise verbally (Hadfield and Koh, 2025). What is new is the delegation market that sets the error-covariance structure of a production economy with closed-form D and $1/M$ loadings, the exact efficiency benchmark and its three failure modes with closed-form wedges, the compute market linking inference to automation, and the factorial agentic-market experiments tied to the theory. Section 2 gives the full audit, including what is not new.

What we do not claim. The efficiency benchmark is a second-order result in a price-taking economy; the information parameters are illustrative and swept rather than estimated; Model C outcomes are simulated bounded-rationality outcomes, never general equilibria; LLM agents are not evidence about human preferences; and all effects are identified within the structural model. Section 14 spells each of these out.

The paper proceeds as follows. Section 2 audits the literature. Section 3 lays out the environment and the equilibrium concepts. Section 4 contains the propositions. Sections 5 and 6 describe Models B and C; Sections 7, 8 and 9 describe computation, the LLM laboratory and the calibration. Sections 10–12 report results, Section 13 robustness, Section 14 limitations, and Section 15 concludes. Appendices contain proofs, algorithms, prompts and schemas, data provenance, replication instructions and additional results.

2. Related literature

AI and the macroeconomy. The macroeconomics of artificial intelligence has been built almost entirely on the task framework. Zeira (1998) and Acemoglu and Restrepo (2018) model automation as capital replacing labor in a measure of tasks, with displacement and reinstatement effects on wages and the labor share; Acemoglu and Restrepo (2022) show that task displacement accounts for most of the rise in U.S. wage inequality, and Acemoglu (2025) uses the same accounting to bound the productivity gains from AI and to allow for "bad" tasks. Aghion et al. (2019) place AI in production and in idea production with Baumol-type bottlenecks, and Jones (2024, 2025) weigh growth against catastrophic risk. Two recent papers make compute the binding constraint: Korinek and Suh (2024) obtain a wage collapse once compute overtakes the task-complexity frontier, and Restrepo (2025) shows that with abundant compute all bottleneck work is automated and wages converge to the cost of replicating human work. Moll et al. (2022) embed automation in a heterogeneous-agent economy and show that it raises returns to wealth, and hence wealth inequality, even when wages rise. Ide and Talamàs (2025) come closest to AI as a decision-maker: in a knowledge-hierarchy economy, autonomous and non-autonomous AI reshape organizations and wages, but errors are not correlated across users and there is no model count. Agrawal et al. (2019) interpret AI as cheaper prediction complementing human judgment, and Athey et al. (2020) characterize the partial-equilibrium allocation of decision rights between a human and an AI with misaligned objectives. On the verbal side, Hadfield and Koh (2025) argue that general-equilibrium distortions from AI consumers and producers are precluded only if markets are large and AI mistakes are zero-mean and independent, and note that copying the same agent across firms induces correlated mistakes that do not wash out; Rothschild et al. (2025), Catalini et al. (2026) and Immorlica et al. (2024) discuss agentic matching, AGI pricing and research agendas. Measurement of exposure, adoption and productivity gains (Eloundou et al., 2024; Bick et al., 2024; Brynjolfsson et al., 2025; Eisfeldt et al., 2023) disciplines our calibration. All of these treat AI as a production input, an information service or a growth factor. We keep the task-based production block of Acemoglu and Restrepo (2018) and the compute constraint of Restrepo (2025), but add a third role: AI as a delegated decision-maker whose forecast error is shared by every principal using the same foundation model. The premise is the one Hadfield and Koh (2025) state in words; our contribution is to make the delegation share and the number of independent models equilibrium objects and to derive their price, welfare and policy consequences.

Dispersed information, common signals and sentiment-driven fluctuations. The information structure of our baseline is classical. Grossman and Stiglitz (1980), Hellwig (1980) and Diamond and Verrecchia (1981) establish that prices aggregate dispersed, independently-noisy signals and cannot be fully revealing when information is costly (see Radner, 1979; Vives, 2008, for the

full-revelation benchmark and a survey). Morris and Shin (2002) show that agents over-weight a common public signal and that more precise public information can lower welfare; Angeletos and Pavan (2007) characterize efficient information use in a general class of economies in which the social value of common information depends on the degree of strategic complementarity. On the business-cycle side, Lorenzoni (2009) shows that noise in a common public signal generates demand-driven fluctuations, Angeletos and La'O (2013) obtain unique-equilibrium fluctuations from exogenously correlated belief errors, and Benhabib et al. (2015) construct self-fulfilling sentiment equilibria; David et al. (2016) quantify the misallocation cost of firm forecast errors, and Sims (2003) microfounds noisy responses through limited attention. We state candidly that "common noise in beliefs is an aggregate shock with real effects" belongs to this literature. What we add is a market in which the correlation of forecast errors is chosen. The information-choice literature has endogenized the source of information: Veldkamp (2006) shows that fixed costs make many agents buy the same news, Hellwig and Veldkamp (2009) derive homogenization of information sources from strategic complementarity, Colombo et al. (2014) and Vives (2017) analyze the welfare properties of endogenous information, and Farboodi and Veldkamp (2020) make the choice of an information *technology* an equilibrium object that changes what prices reveal (see also Farboodi et al., 2022; Farboodi and Veldkamp, 2023; Veldkamp, 2011). Jones and Tonetti (2020) show that ownership of a non-rival information input shapes its use, and Dugast and Foucault (2018) that cheap raw data can lower price informativeness without any correlated-error story. In each of these papers the technology is private, or homogenization is driven by complementarity or fixed costs. Our delegation margin instead rests on a private precision gain and the scale economies of foundation models, with no strategic complementarity required: a household or firm that delegates obtains model m's forecast and inherits model m's error, so the error covariance of the economy loads on the delegation share D and on $1/M$. The efficiency analysis then connects directly to Angeletos and Pavan (2007): in the constant-returns, price-taking benchmark the private and social marginal values of delegation coincide exactly for every D, M and degree of strategic substitutability (Proposition 3), so aggregate "model noise" is real but not an inefficiency, and a wedge opens only under identified frictions—non-pecuniary aggregate effects, uninsurable price risk for non-participants, prices in collateral constraints, or finite market depth (Proposition 4).

Algorithmic trading, homogenization and AI in finance. That algorithmic behavior is more correlated than human behavior is documented: Chaboud et al. (2014) find that algorithmic strategies in foreign-exchange markets are more correlated than human strategies yet improve efficiency, Khandani and Lo (2011) show that similar, crowded quantitative strategies unwound together in August 2007, and Kirilenko et al. (2017) document the synchronized withdrawal of high-frequency liquidity during the Flash Crash. The theory of crowding is also established: Stein (2009) analyzes

crowded-trade and leverage externalities in which each arbitrageur ignores the effect of his trade on others' unwinding, and Wagner (2011) shows that diversification by every investor makes portfolios similar, creates joint-liquidation risk and leaves diversity under-provided. Outside finance, Kleinberg and Raghavan (2021) show that a shared, more accurate algorithm can lower aggregate outcome quality because independence is lost, Bommasani et al. (2022) document outcome homogenization across systems sharing a base model, Doshi and Hauser (2024) find that generative assistance raises individual quality while reducing collective diversity, and Jo et al. (2025) show that correlated predictors soften product-market competition. Calvano et al. (2020) and Fish et al. (2024) obtain correlated algorithmic behavior through learned collusion rather than shared errors. In asset markets, Dou et al. (2025) show that reinforcement-learning traders in a Kyle-type setting (Kyle, 1985) sustain supra-competitive outcomes through price-trigger strategies and through homogenized learning biases that lower informativeness and liquidity; Park (2026) analyzes correlated AI signals in a noisy rational-expectations market; Qiu and Han (2026) simulate how representation homogeneity compresses forecast disagreement under stress; and Gârleanu and Pedersen (2018) show that delegation to human managers already co-determines market efficiency. Gabaix and Koijen (2021) provide the transmission channel we require: in inelastic markets, correlated flows move prices. The closest formal antecedent is Meng and Chen (2026), who study endogenous adoption of AI trading with a common signal component in a Kyle market, obtain a systemic-risk multiplier convex in the adoption share, and show tipping to monoculture in a supermodular adoption game. We differ on four counts: our adopters are households and firms whose delegation moves wages, the interest rate and capital accumulation, not traders alone; the model count M enters explicitly, so diversity is a comparative static; we derive an efficiency benchmark and the frictions under which a Pigouvian delegation tax or model diversity restores constrained efficiency; and we close the compute market. Relative to Stein (2009) and Wagner (2011), our externality operates in forecast space rather than portfolio space and is priced through factor and asset prices rather than liquidation costs. The 2025–26 evidence constrains rather than confirms us. He (2026) uses ChatGPT outages to show that the tool acts as a belief homogenizer for retail investors, synchronizing trades and worsening liquidity (on the tradability of a single public model's signal, see Lopez-Lira and Tang, 2023); Spiro (2026) finds that LLM forecast errors are strongly correlated across vendors, which is why we treat M as an *effective* number of independent models to be calibrated rather than a count of providers; and Hansen and Lee (2025) find that LLM agents herd less than humans in observational cascade games, consistent with our channel, which runs through shared technology rather than observation (contrast Banerjee, 1992). Policy institutions have stated the monoculture concern verbally (Gensler and Bailey, 2020; Daníelsson et al., 2022; Danielsson and Uthemann, 2024; Financial Stability Board, 2024; International Monetary Fund, 2024; Bank for International Settlements, 2024). We also confront the competing explanation. Belief homogenization reduces

disagreement, and in Miller (1977), Harrison and Kreps (1978), Scheinkman and Xiong (2003) and Simsek (2013), surveyed by Hong and Stein (2007), less disagreement means less overpricing, less speculative volume and less leverage. Our Model C reproduces this in calm periods; the model-noise channel dominates only when the shared error is large relative to the dispersion it removes, and we characterize when each effect wins.

Collateral, leverage and pecuniary externalities. The amplification block is standard and we claim nothing new about it. Kiyotaki and Moore (1997) and Bernanke et al. (1999) establish the feedback between asset prices, net worth and borrowing; Brunnermeier and Pedersen (2009) the funding–market liquidity spiral; Geanakoplos (2010) the leverage cycle; Gertler and Kiyotaki (2010) the bank balance-sheet channel; and Brunnermeier and Sannikov (2014) the non-linear dynamics of continuous-time macro-finance. The welfare logic is that of Greenwald and Stiglitz (1986): with incomplete markets, equilibria are generically constrained inefficient because of pecuniary externalities. Lorenzoni (2008) shows over-borrowing when agents ignore the price effects of fire sales, and Dávila and Korinek (2018) decompose pecuniary externalities into distributive and collateral components with the corresponding Pigouvian corrections. Our collateral wedge is a Dávila–Korinek collateral externality whose triggering shock is model noise: a shared forecast error moves collateral values for every delegator at once, and the private delegation decision ignores the resulting tightening of others' constraints. We note the tension with Geanakoplos (2010) and Simsek (2013), whose leverage cycles require belief heterogeneity that delegation reduces; this is the calm-period effect above.

Delegation and principal–agent theory. We model an AI agent as a constrained policy acting under a mandate, with the principal retaining residual claims and liabilities. This is a contract on observables in the sense of Holmstrom (1979), whose informativeness principle governs what the mandate conditions on; the linear risk limits we use follow Holmstrom and Milgrom (1987), and multi-task mandates covering both trading and risk control distort effort as in Holmstrom and Milgrom (1991). The delegation decision itself is an Aghion and Tirole (1997) trade-off between initiative and control, with the threshold structure of Dessein (2002) and the interval-delegation sets of Alonso and Matouschek (2008). What distinguishes our delegation threshold (Proposition 2) from these partial-equilibrium results, and from Athey et al. (2020), is the general-equilibrium term attached to it: the mandate's payoff depends on how many other principals have delegated to the same model. The robo-advice evidence supports the private side of the trade-off: D'Acunto et al. (2019) find that adopters diversify more and that under-diversified investors gain, and Rossi and Utkus (2024) estimate welfare gains. Our model predicts that such users' trades become correlated, a testable implication.

Heterogeneous agents, agent-based economics and LLM simulations. Model B is a stationary recursive equilibrium in the tradition of Huggett (1993) and Aiyagari (1994), with aggregate risk handled as in Krusell and Smith (1998); we solve it with the sequence-space methods of Auclert et al. (2021) and relate the continuum-with-common-noise structure to the mean-field formulation of Achdou et al. (2022). The direct-versus-indirect decomposition of Kaplan et al. (2018) organizes our split of delegation effects into a private precision gain and a general-equilibrium model-noise channel. Because Hubmer et al. (2021) and Moll et al. (2022) already explain rising top wealth shares through heterogeneous returns and automation-driven increases in the return to wealth, the ownership result (Proposition 6) must carry a distinct signature: compute scarcity rents respond to the delegation share while the generic return to capital does not. Model C is a bounded-rationality simulated outcome in the constructive sense of Tesfatsion (2006), benchmarked against the stylized facts of agent-based finance (LeBaron, 2006; Hommes, 2006) and macro (Farmer and Foley, 2009; Dawid and Delli Gatti, 2018; Axtell and Farmer, 2025). Brock and Hommes (1998) already endogenize predictor choice by past fitness; ours is predictor choice with a shared predictor and a price externality. The generative-agent literature supplies the architecture (Park et al., 2023) and the caution of Horton et al. (2023); Chen et al. (2023) and Mei et al. (2024) show that LLM choices are broadly rational and within the human distribution, Ross et al. (2024) document systematic, model-specific biases—exactly the model-indexed error we posit—and Bybee (2023) shows that LLM expectations track survey expectations. LLM-agent simulations exist: Li et al. (2024) reproduce the Phillips curve and Okun's law with LLM households, Zhao et al. (2023) study competing LLM firms, Lopez-Lira (2025) runs LLM traders in an order book and warns that similar architectures may behave in correlated ways, Ross et al. (2026) find that higher LLM participation improves price discovery under a "better models, riskier systems" framing, Hammond et al. (2025) list correlated failures from shared models as a multi-agent risk, and Xia et al. (2026) flag strategy homogenization as an open problem. None varies the model count or the adoption margin as a treatment tied to a theory. Our laboratory follows the discipline of Gode and Sunder (1993): constraints, budgets, collateral and accounting are enforced in deterministic code, zero-intelligence and Bayesian controls isolate what the agent contributes, and LLM output is never trusted as accounting truth or read as evidence on human preferences.

Summary of the gap. The ingredients of our mechanism are all in the literature: common noise generates fluctuations, common information is over-used, information sources homogenize endogenously, private similarity creates inefficient common exposure, pecuniary externalities justify Pigouvian corrections, and collateral constraints amplify shocks. Four things are not. First, no existing model makes the *correlation structure* of forecast errors—not only their variance—an equilibrium object chosen by households and firms through a delegation decision, with a closed-

form loading on D and $1/M$ (Proposition 1). Second, no paper states the delegation externality in a production economy: competitive delegation is constrained efficient in the constant-returns benchmark and becomes excessive only when model noise is loaded on wages, interest rates and asset prices through identified frictions, with a Pigouvian delegation tax and model diversity as equivalent remedies derived with the Dávila and Korinek (2018) decomposition (Propositions 3 and 4). Third, a compute market that clears inference demand from delegation against automation demand from production through a single rental price, generating rents to the owners of algorithmic capital, is a new mechanism with a comparative static that distinguishes it from a generic rise in the return to capital (Propositions 5 and 6). Fourth, our agentic market experiments vary M and D factorially with LLM agents across Walrasian, continuous-double-auction and bilateral designs, and tie each treatment to a prediction of the theory.

3. Institutions and environment

This section lays out the baseline economy (Model A). Two later sections add idiosyncratic risk and heterogeneous ownership (Model B, Section 5) and an explicit asset market with credit (Model C, Section 6). Throughout, we keep four roles of artificial intelligence apart (Table 1): a production technology, an information-processing service, a delegated decision-maker, and a hypothetical independent economic principal. The first three exist today and have identified legal and economic counterparties; the fourth is a counterfactual that we label as such wherever it appears.

Table 1. Four roles of AI and where each enters the model

Role	Object in the model	Residual claimant and liability
(i) Production technology	compute Q^P performs tasks $x \leq \iota$ in the task aggregator (2)	the firm; compute owners receive the rental p_Q
(ii) Information service (advisory)	the human sees her own signal and the model's signal and decides herself	the household
(iii) Delegated decision-maker (autonomous)	the model's policy π_m applies the household's decision rule to the model's forecast	the household (principal); the fee ϕ is paid to compute and energy suppliers
(iv) Independent AI principal	an agent with its own balance sheet and no human residual claimant (Model C only)	none; a labelled counterfactual

3.1. Agents, goods and shocks

Time is discrete, $t = 0, 1, 2, \ldots$, and there is one final good, the numéraire. The economy is populated by a continuum of households $i \in [0,1]$, a representative competitive final-good firm, a competitive compute-producing sector, an energy sector with constant marginal cost p_E (goods per unit of energy), and M foundation models $m = 1, \ldots, M$, which are policies rather than agents. Aggregate log-productivity follows $z_t = \rho_z z_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ i.i.d.

An AI agent is a constrained policy. A model-m agent is a map $\pi_m(x \mid o, \mathscr{M}_m, \text{mandate}, \varepsilon)$ from an observation o, the model's information $\mathscr{M}_m$, a mandate (the principal's objective and constraints) and randomness ε into an action x. It has no utility, no balance sheet and no liability. In Models A and B the mandate is "maximise my expected utility given your information", so π_m is the household's own optimal decision rule evaluated at the model's forecast. In Model C the policy is either that rule (a Bayesian control), a rule-based policy, or a live language model whose proposed action is validated against a schema and clipped by deterministic constraints before it touches any ledger.

3.2. Information

The only friction of Model A is informational and one step ahead. After z_t is known, decision-makers receive news about $\theta_t \equiv \varepsilon_{t+1}$, the next innovation. A human receives a private signal $s_{it} = \theta_t + e_{it}$, $e_{it} \sim N(0, \tau_{H,i}^{-1})$, independent across households and time. Model m produces a signal $s_{mt} = \theta_t + \eta_{mt}$, $\eta_{mt} \sim N(0, \tau_A^{-1})$, independent across models and time but *common to every principal that delegates to model m*. The prior precision is $\tau_\theta = \sigma_\varepsilon^{-2}$ and the Bayesian weights are $\beta_H = \tau_H/(\tau_\theta + \tau_H)$ and $\beta_A = \tau_A/(\tau_\theta + \tau_A)$. All uncertainty about θ_t is resolved at $t+1$, so the economy returns to a full-information recursive structure every period. The assumption that model errors are common is the formal counterpart of the observation that users of one foundation model share its "world view" at a point in time; the empirical literature reviewed in Section 2 finds cross-user and even cross-vendor error correlations far above zero, so M should be read as an *effective* number of independent models.

3.3. Technology

Tasks $x \in [0,1]$ are combined with unit elasticity into a task bundle T_t, which is combined with capital:

$$Y_t = e^{z_t} K_t^{\alpha_K} T_t^{1-\alpha_K}, \qquad \ln T_t = \int_0^1 \ln y_t(x)\,\mathrm{d}x. \tag{1}$$

Task x is produced either by labour, $y(x) = l(x)$, or—if $x \le \bar{\iota}$, the technological frontier—by compute, $y(x) = \psi(x)q(x)$ with $\psi(x) = \psi_0 e^{-\psi_1 x}$ decreasing in the task index (easy tasks are automated first). Cost minimisation automates $x \le \iota_t = \min\{\bar{\iota}, \iota_t^*\}$, where ι_t^* solves $p_{Q,t}/\psi(\iota_t^*) = w_t$. With unit elasticity across tasks every automated task uses Q_t^P/ι_t units of compute and every manual task $L_t/(1-\iota_t)$ units of labour, so

$$T_t = \Psi(\iota_t)\left(\frac{Q_t^P}{\iota_t}\right)^{\iota_t}\left(\frac{L_t}{1-\iota_t}\right)^{1-\iota_t}, \qquad \Psi(\iota) = \exp\left(\int_0^{\iota} \ln\psi(x)\,\mathrm{d}x\right) = \psi_0^{\iota} e^{-\psi_1 \iota^2/2}. \tag{2}$$

Competitive factor prices are $R_t^K = \alpha_K Y_t/K_t$, $w_t = (1-\alpha_K)(1-\iota_t)Y_t/L_t$ and $p_{Q,t} = (1-\alpha_K)\iota_t Y_t/Q_t^P$. Combining the automation threshold with the two last expressions gives the interior condition

$$\psi(\iota_t) = \frac{p_{Q,t}}{w_t} = \frac{\iota_t}{1-\iota_t}\frac{L_t}{Q_t^P}, \tag{3}$$

whose left side decreases and right side increases in ι (Lemma 1). Factor shares are α_K for capital, $(1-\alpha_K)\iota_t$ for compute and $(1-\alpha_K)(1-\iota_t)$ for labour. This is the standard task-based structure of Acemoglu and Restrepo (2018); it is not the contribution of the paper, but it is needed so that the decision channel and the production channel compete for the same compute.

3.4. Compute

Compute is a stock $K_{Q,t}$, predetermined at t, rented at $p_{Q,t}$ per unit and period. It has two uses: production (Q_t^P) and inference for delegated decisions, $\bar{q}D_t$, where $\bar{q}$ is the compute consumed by one delegated household-decision per period and $D_t \in [0,1]$ is the delegation share. The compute market clears,

$$Q_t^P + \bar{q}D_t = K_{Q,t}, \qquad p_{Q,t} \ge 0. \tag{4}$$

The stock accumulates according to $K_{Q,t+1} = (1-\delta_Q)K_{Q,t} + I_{Q,t}/p_{Q,t}^K$, where $p_{Q,t}^K = p_{Q,0}^K(1-g_Q)^t$ is the (declining) price of compute capital in goods. Compute owners are households with shares ω_i, $\int \omega_i\,\mathrm{d}i = 1$; they receive $\Pi_{Q,t} = p_{Q,t}K_{Q,t}$ and finance $I_{Q,t}$. In the short run $K_{Q,t}$ is fixed and $p_{Q,t}$ carries scarcity rents; in the long run, with elastic supply, p_Q converges to the user cost $u_{Q,t} \approx (r_t + \delta_Q - g_Q)p_{Q,t}^K$. Model B interpolates with a supply elasticity ε_Q.

3.5. Households

Preferences are $\mathbb{E}_0\sum_t \beta^t\left[u(c_{it}) - v(l_{it}) - \kappa_i d_{it}\right]$ with $u(c) = c^{1-\gamma}/(1-\gamma)$, $v(l) = \chi l^{1+1/\nu}/(1+1/\nu)$, and $\kappa_i \ge 0$ a delegation disutility (trust, agency loss) distributed with c.d.f. F_κ. The delegation indicator $d_{it} \in \{0,1\}$ is chosen at the start of the period, before signals arrive. Time is divided

between work and supervision: $l_{it} + \tau_s d_{it} \leq 1$. The budget constraint is

$$c_{it} + a_{i,t+1} + \phi_t d_{it} = w_t l_{it} + R_t a_{it} + \omega_i \big(\Pi_{Q,t} - I_{Q,t}\big), \qquad R_t = 1 - \delta + R_t^K, \tag{5}$$

with the delegation fee $\phi_t = p_{Q,t}\bar{q} + p_E \bar{e}$ (compute rental paid to compute owners, energy paid to the energy sector). A borrowing limit $a_{i,t+1} \geq \underline{a}$ is slack in Model A and binds for some households in Model B. The transversality condition is $\lim_T \beta^T \mathbb{E}[u'(c_{iT}) a_{i,T+1}] = 0$.

Decision rule and delegation modes. Let $a^*(\cdot)$ denote the household's optimal savings rule under its information. A self-managing household applies it at $\mathbb{E}[\theta_t \mid s_{it}]$; a delegating household's agent applies the same rule at $\mathbb{E}[\theta_t \mid s_{mt}]$ (the mandate is the principal's objective). Advisory AI (role ii) applies it at $\mathbb{E}[\theta_t \mid s_{it}, s_{mt}]$ at the cost of a larger supervision time $\tau_s^{adv} > \tau_s$. The difference between modes is therefore the weight on the common error, not the decision rule itself.

3.6. Market clearing, flow of funds and the equilibrium concept

Labour, capital, compute and goods markets clear:

$$L_t = \int l_{it}\,\mathrm{d}i, \quad K_{t+1} = \int a_{i,t+1}\,\mathrm{d}i, \quad (4), \quad C_t + I_t + I_{Q,t} + p_E \bar{e} D_t = Y_t, \quad I_t = K_{t+1} - (1-\delta)K_t. \tag{6}$$

Every flow has a payer and a payee: fees $p_Q \bar{q} D$ go from delegating households to compute owners; energy $p_E \bar{e} D$ goes to the energy sector and is a resource use; factor payments flow from the firm to households; compute investment flows from owners to the compute-producing sector. These identities are asserted in the code after every operation (Appendix E).

Definition 1 (Recursive competitive equilibrium with dispersed information)**.** Let $S_t = (K_t, K_{Q,t}, z_t)$. An equilibrium consists of price functions (w, R^K, p_Q), household policies (c, a', l, d) measurable with respect to each household's information, firm policies, an aggregate law of motion $K_{t+1} = \mathscr{K}(S_t, \theta_t, \eta_{1t}, \ldots, \eta_{Mt})$ and a delegation share $D(S_t)$ such that (a) policies are optimal given beliefs; (b) beliefs are Bayesian given signals and the law of motion; (c) the markets in (4) and (6) clear; (d) $D(S_t) = \int d_{it}\,\mathrm{d}i$ is consistent with individual delegation choices.

Definition 2 (LQ-recursive equilibrium)**.** The second-order approximation of Definition 1 around the deterministic steady state, in which the within-period decision is the linear-quadratic game of Section 4 and the between-period dynamics are the log-linearised economy with the aggregate action of the game entering the capital law of motion.

Propositions are stated for Definition 2; the numerical sections check their accuracy. In the *zero-AI limit* ($K_Q = 0$, hence $\iota = 0$ and $D = 0$) equations (1)–(6) collapse to the stochastic growth

model $Y = e^z K^{\alpha_K} L^{1-\alpha_K}$, which the code reproduces against the Brock–Mirman closed form when $\delta = 1$ and $\gamma = 1$ (Appendix B).

4. Benchmark equilibrium: the delegation game

This section contains the analytical core. We first reduce the within-period decision problem to a linear-quadratic game with dispersed information, then characterise delegation, aggregate "model noise", and the private and social values of delegation. Proofs are in Appendix A; every closed form below is also implemented in `age.equilibrium.lq_game` and cross-checked against Monte Carlo simulations and against a direct computation of expected utility in the test suite.

4.1. From the economy to a quadratic game

Let x_i denote household i's log-deviation of its payoff-relevant action from the full-information steady-state path and $X = \int x_i \mathrm{d}i$ the aggregate action. Expanding the household's objective to second order around the symmetric steady state gives

$$U(x_i, X, \theta) = \tfrac{1}{2} U_{xx} x_i^2 + U_{xX} x_i X + \tfrac{1}{2} U_{XX} X^2 + U_{x\theta} x_i \theta + U_{X\theta} X \theta + \ldots, \tag{7}$$

where the omitted terms do not involve x_i. Normalising by $|U_{xx}|$, the best response is $x_i = \mathbb{E}_i[\kappa_0 \theta + \alpha X]$ with $\kappa_0 = -U_{x\theta}/U_{xx}$ (the partial-equilibrium response to a known θ) and $\alpha = -U_{xX}/U_{xx}$ (the degree of strategic complementarity). In capital accumulation $U_{xX} = \beta u'(c') F_{KK} < 0$: actions are strategic substitutes, $\alpha < 0$. The class is that of Angeletos and Pavan (2007); what is new is the information structure, which is chosen in a market.

Two mappings from the economy of Section 3 to (7) are used. In the *savings mapping*, x_i is the log deviation of savings $a_{i,t+1}$ and θ the TFP news; κ_0 and α are computed from the log-linearised household problem (Appendix B). In the *portfolio mapping*, households allocate wealth between risky capital claims and a safe store of value with mean-variance preferences; θ is the predictable component of next period's return,

$$\kappa_0 = \frac{1}{\gamma \sigma_R^2}, \qquad \alpha = -(1 - \alpha_K) \bar{R}^K \kappa_0, \qquad |U_{xx}| = \gamma \sigma_R^2 \bar{K} \text{ (goods)}, \tag{8}$$

where σ_R is the return standard deviation and α_K the capital share. Section 10 reports that the savings mapping yields a negligible stake ($\kappa_0 \approx -0.0015$): wealth and substitution effects of productivity news on saving nearly cancel. The portfolio mapping is therefore our baseline. The lesson is itself a result: *which* decision is delegated determines whether delegation has aggregate consequences.

Lemma 1 (Automation threshold)**.** *For $Q^P > 0$ equation (3) has a unique solution $\iota^*(Q^P/L) \in (0,1)$, increasing in Q^P/L; the automated share is $\iota = \min\{\bar{\iota}, \iota^*\}$, and factor payments exhaust output.*

Lemma 2 (Constant-returns restriction)**.** *With constant returns, price-taking and symmetric ownership, the marginal effect of the aggregate action on an individual's payoff vanishes at symmetric allocations for every state, $U_X \equiv 0$; hence $U_{X\theta} = 0$ and $U_{xX} + U_{XX} = 0$. The social curvature of welfare in the aggregate action is $(1-\alpha)|U_{xx}|$.*

The lemma is Euler's theorem applied to factor payments: a change in aggregate capital redistributes income between capital, labour and compute, but the first-order sum of these transfers is zero when each household receives all three incomes in the same proportions. We write $U_{XX} = -\upsilon U_{xX}$ in general and call $\upsilon = 1$ the constant-returns benchmark.

4.2. Linear equilibrium and aggregate model noise

A share D of households delegates, split evenly across the M models. We look for equilibria in which non-delegators play $x_i = a_H s_i$ and the users of model m play $x_i = a_A s_m$.

Proposition 1 (Linear equilibrium)**.** *For every $\alpha < 1$, $D \in [0,1]$ and $M \geq 1$ there is a unique linear equilibrium,*

$$a_H = \beta_H c, \qquad a_A = \beta_A \xi c, \qquad \xi = \frac{1}{1 - \alpha\frac{D}{M}(1-\beta_A)}, \qquad c = \frac{\kappa_0}{1 - \alpha\big[(1-D)\beta_H + D\beta_A\xi\big]}, \tag{9}$$

and the aggregate action is

$$X = \bar{A}\theta + \frac{Da_A}{M}\sum_{m=1}^{M}\eta_m, \qquad \bar{A} = (1-D)a_H + Da_A, \qquad \mathrm{Var}\left(\frac{Da_A}{M}\sum_m \eta_m\right) = \frac{D^2 a_A^2}{M\tau_A}. \tag{10}$$

When $\alpha < 0$, $\xi < 1$: a delegator, knowing that a mass D/M of others act on the same error and that actions are substitutes, dampens its response. The law of large numbers removes the non-delegators' errors from X and leaves the delegators' errors in it with weight D and variance $\propto D^2/M$.

Two aggregate consequences follow. First, X tracks the fundamental more closely as D rises ($\bar{A}$ moves toward the full-information response $\kappa_0/(1-\alpha)$ because the model is better informed than the median human). Second, X acquires a component that has nothing to do with fundamentals. Wages, the rental rate and asset prices inherit this component, which we call *model noise*. Figure 1(b) plots its variance.

4.3. Private value of delegation and the adoption threshold

A measure-zero household compares the precision of its target $\hat{x}_i = \kappa_0\theta + \alpha X$ under its own signal and under the model's signal.

Proposition 2 (Private value and adoption threshold)**.** *The private value of delegation (gross of fees, in units of $|U_{xx}|$) is*

$$\mathrm{PV}(D) = \tfrac{1}{2}\big[\mathrm{Var}(\hat{x}_i \mid s_i) - \mathrm{Var}(\hat{x}_i \mid s_m)\big], \tag{11}$$

which at $\alpha = 0$ reduces to $\mathrm{PV} = \frac{\kappa_0^2}{2}\frac{\tau_A - \tau_H}{(\tau_A+\tau_\theta)(\tau_H+\tau_\theta)}$; it is increasing in τ_A, decreasing in τ_H, and strictly positive if and only if $\tau_A > \tau_H$. Household i delegates if and only if

$$\mathrm{PV}(D)\,|U_{xx}| \;\geq\; u'(\bar{c})\phi + \kappa_i + \tau_s \bar{w} u'(\bar{c}), \tag{12}$$

so the delegation share solves the fixed point $D = F_\kappa\big(\mathrm{PV}(D)|U_{xx}| - u'(\bar{c})\phi - \tau_s\bar{w}u'(\bar{c})\big)$. A solution exists; it is increasing in τ_A and decreasing in the fee ϕ and in supervision time τ_s.

The value of the shared signal is not constant in D: it includes a "knowing what others know" component in the sense of Hellwig and Veldkamp (2009), because the model's error is also informative about the aggregate action. With substitutes this component makes PV *decrease* in D (Figure 1(a)): the more people act on the same model, the less a marginal user gains, and a model only slightly better than the median human can be worth less than nothing once widely adopted. Adoption is self-limiting without any congestion in compute.

4.4. The efficiency benchmark

Welfare per capita, relative to full information and in units of $|U_{xx}|$, is $W(D) = \mathbb{E}\,U(x_i, X, \theta) - D\tilde{\phi}$, which under Lemma 2 decomposes as

$$W(D) = -\tfrac{1}{2}\Big[(1-\alpha)\,\mathbb{E}(X - X^*)^2 + \mathbb{E}(x_i - X)^2\Big] - D\tilde{\phi}, \qquad X^* = \frac{\kappa_0}{1-\alpha}\theta, \tag{13}$$

the sum of an aggregate-volatility loss and a dispersion loss (Angeletos and Pavan, 2007).

Proposition 3 (Efficiency benchmark)**.** *Under Lemma 2 ($\upsilon = 1$, $U_{X\theta} = 0$), for every $D \in [0,1]$, $M \geq 1$, $\alpha < 1$ and precisions $(\tau_\theta, \tau_H, \tau_A)$,*

$$\mathrm{PV}(D) = \frac{\partial}{\partial D}\big[W(D) + D\tilde{\phi}\big]. \tag{14}$$

The competitive delegation share is constrained efficient: the planner who can choose D (but cannot pool signals) chooses the same D as the market, for any number of models M.

This is the first organising result of the paper, and it is a negative one for the most common policy narrative. Model noise is real—wages, returns and capital fluctuate with the foundation model's mistakes—but in a price-taking economy with constant returns and complete markets it is not an inefficiency. Delegators already dampen their response to the common signal exactly as a planner would (strategic substitutability), and the pecuniary effects of their common error on everyone else are zero-sum. The result is the counterpart, for the *acquisition* of a correlated signal, of the efficiency of the use of information in neoclassical economies (Angeletos and Pavan, 2007; Colombo et al., 2014); it was derived symbolically before any numerical work, and the identity "wedge $= 0$" is a property-based test of the replication package.

4.5. When delegation becomes a policy problem

Proposition 4 (Sources of a delegation externality)**.** *Let* $\Delta(D) \equiv \mathrm{PV}(D) - \partial[W + D\tilde{\phi}]/\partial D$ *denote the wedge between the private and social marginal value of delegation (positive: over-delegation).*

(a) Non-pecuniary aggregate effects. *With* $U_{XX} = -\upsilon U_{xX}$ *and* $U_{X\theta} = 0$, $\Delta(D) = (\upsilon - 1)\,\frac{\alpha}{2}\,\frac{\partial \mathbb{E}[X^2]}{\partial D}$. *With strategic substitutes (*$\alpha < 0$*) and* $\partial \mathbb{E}[X^2]/\partial D > 0$*, there is over-delegation if* $\upsilon < 1$ *and under-delegation if* $\upsilon > 1$.

(b) Uninsurable price risk. *If a mass* m_w *of hand-to-mouth workers consume their wage and the log wage loads on the aggregate action with elasticity* α_K*, then* $\Delta_b(D) = m_w\,\frac{1}{2}\gamma\bar{w}\alpha_K^2\,\partial\,\mathrm{Var}(X)/\partial D > 0$ *in goods per unit of D, and the Pigouvian delegation tax equals* Δ_b*. Since* $\partial\,\mathrm{Var}(X)/\partial D$ *contains the model-noise term* $2Da_A^2/(M\tau_A)$*, the tax falls with model diversity at rate* $1/M$.

(c) Collateral constraints. *When prices enter borrowing constraints, the price impact of correlated orders creates a pecuniary externality of the Dávila and Korinek (2018) type (Proposition 7).*

In the benchmark of Proposition 3 model diversity M is welfare-irrelevant; under (b) and (c) it is a substitute for the tax.

4.6. Compute market and ownership

Proposition 5 (Compute market in the short run)**.** *With* K_Q *fixed, equations (3), (4) and the threshold rule (12) have a unique equilibrium* (p_Q, Q^P, D)*. An improvement in model quality (higher* τ_A*) raises D and* p_Q *and lowers* Q^P *and* ι*; the labour share* $(1-\alpha_K)(1-\iota)$ *rises. With compute supplied elastically at user cost,* Q^P *is unaffected by delegation and* ι *and D move together.*

Proposition 6 (Ownership of algorithmic capital)**.** *Compute income* $(1-\alpha_K)\iota Y + p_Q\bar{q}D$ *accrues to households in proportion to* ω_i*. In a two-class economy in which a mass n of owners holds all compute and a pro-rata share of physical capital, the owners' income share is* $n\alpha_K + (1-\alpha_K)\iota$ *plus scarcity rents; it increases with automation* ι *and with the rents, while broad ownership (*$\omega_i = 1$*) leaves the income distribution unchanged by* ι *at given factor prices.*

4.7. Collateral amplification

Proposition 7 (Margin spiral on model noise)**.** *Consider a mass D of leveraged delegators holding h units each at the margin limit $\ell = 1-m$, a liquidity provider with relative price impact λ_r per unit sold, and a common negative model error that moves the price by $\varepsilon_0 = \lambda_r D c\,\bar{\eta} < 0$ on impact. Forced sales to restore the margin constraint multiply the price change,*

$$\varepsilon = \mu\,\varepsilon_0, \qquad \mu = \frac{1}{1-\lambda_r h D\,\frac{1-m}{m}}, \qquad \mathrm{Var}(\varepsilon) = \mu^2 \lambda_r^2 D^2 c^2 \frac{\sigma_A^2}{M}, \tag{15}$$

provided $\lambda_r h D(1-m)/m < 1$; the multiplier increases in leverage $(1-m)/m$, in the mass of leveraged delegators D and in price impact, and the spiral is unbounded as $\lambda_r h D(1-m)/m \to 1$. Model diversity lowers the variance at rate $1/M$ but leaves the multiplier unchanged.

The proposition is static and deliberately minimal; Model C (Section 6) checks it with explicit balance sheets, defaults and three market designs.

5. Heterogeneous agents and ownership

Model A says nothing about who delegates and who owns the compute. Model B is a stationary heterogeneous-agent economy in the tradition of Aiyagari (1994) in which the decision channel of Section 4 appears in the cross-section as a *decision-quality* return term, and compute ownership is either broad or concentrated.

5.1. Households

A continuum of households has CRRA utility $u(c) = c^{1-\gamma}/(1-\gamma)$, discount factor β, and earnings $w\exp(\varepsilon_{it})$ with $\varepsilon' = \rho_e \varepsilon + \sigma_e \nu$ discretised by the Rouwenhorst method on seven states. Savings $a' \geq \underline{a}$ earn the fund return R plus an idiosyncratic term that reflects the quality of the household's own financial decisions:

$$x_{i,t+1} = (R + \nu^r_{i,t+1})\,a_{i,t+1} + w\exp(\varepsilon_{i,t+1}), \qquad \nu^r \sim \begin{cases} N(-m_H, \sigma^2_{rH}) & d=0 \text{ (self-managed)} \\ N(0, \sigma^2_{rA}) & d=1 \text{ (delegated)}, \end{cases} \tag{16}$$

at a fee $\phi_0 + \phi_1 a'$ paid when delegating. The mean cost of mistakes m_H is a resource cost (it is destroyed, like trading costs); the dispersion is a zero-sum redistribution across households (it averages out). The household chooses (c, a', d) every period:

$$V(x,\varepsilon) = \max_{d\in\{0,1\}} \max_{a'\geq \underline{a}} u\big(x - a' - d(\phi_0 + \phi_1 a')\big) + \beta\,\mathbb{E}\big[V(x',\varepsilon') \mid \varepsilon, d\big]. \tag{17}$$

Because the gain from a better decision is proportional to the stake a' while part of the fee is fixed, the rich delegate first: this is Proposition 2 in the cross-section, and it feeds back into accumulation. The specification is deliberately reduced-form—it stands in for the portfolio-choice stake of Section 4 in an economy without aggregate risk—and the size of m_H and σ_{rH} is informed by the household-finance evidence on investment mistakes and return heterogeneity discussed in Section 9.

5.2. Firms, compute and ownership

The firm uses the task technology (1)–(3) with a compute stock supplied along $K_Q = K_{Q,0}(p_Q/u_Q)^{\varepsilon_Q}$ ($\varepsilon_Q = 0$: fixed stock; $\varepsilon_Q \to \infty$: perfectly elastic at user cost u_Q). Inference for delegated decisions uses $\bar{q}D$ units of compute, so the compute market (4) links the adoption share $D = \int d\,\mathrm{d}\lambda$ to production automation. Physical capital and the user-cost part of compute are held through a mutual fund that pays $R = 1 + r$ (plus, under broad ownership, the scarcity rents $(p_Q - u_Q)K_Q$ in proportion to wealth). Under *concentrated* ownership the scarcity rents accrue to a measure-n_K owner class outside the household block, whose members hold the capitalised rent $(p_Q - u_Q)K_Q/(n_K r)$ and are added to the wealth distribution as a point mass—the two-class economy of Proposition 6.

5.3. Equilibrium and computation

Definition 3 (Stationary recursive competitive equilibrium)**.** A value function V, policies (c, a', d), an invariant distribution λ over (x, ε), prices (r, w, p_Q) and quantities (K, K_Q, Q^P, D) such that (i) V solves (17); (ii) λ is invariant under the policies and the shocks; (iii) $K + p_Q^K K_Q = \int a'\,\mathrm{d}\lambda$ (asset market), $L = \int \exp(\varepsilon)\,\mathrm{d}\lambda = 1$, $Q^P + \bar{q}D = K_Q$ with K_Q on its supply curve, and factor prices are the marginal products of Section 3.3.

The household problem is solved by value-function iteration with a discrete delegation choice and Howard acceleration (numba kernels), the distribution by non-stochastic simulation on the cash-on-hand grid, and the interest rate by a scan-and-Brent root search on excess assets (Appendix B). The zero-AI limit is a standard Aiyagari economy and is used as a control; the channel decomposition switches on, one at a time, production compute (productivity and displacement), delegation (information), ownership and the borrowing limit.

6. Asset markets and credit

Model C replaces the Walrasian auctioneer of Models A and B with explicit trading, balance sheets, collateralised credit and defaults. It is where the "agentic" part of the title is taken literally: AI agents place orders, borrow within principal-approved limits and rebalance, and the outcome is a

temporary trading equilibrium per period rather than a rational-expectations equilibrium. We never call these outcomes general equilibria; feasibility and market clearing are verified by assertion, and the Bayesian mean-variance agent is the control against which every LLM result is read.

6.1. Assets, balance sheets and the bank

There are N investors, one bank, one market and a regulator (who sets the margin m and the bank capital ratio k_B). A risky security in fixed supply S pays a dividend $d_t = \bar{d} + \rho_d(d_{t-1} - \bar{d}) + \sigma_d u_t$; a safe deposit/loan at rate r_f is intermediated by the bank. Investor n's balance sheet is deposits c_n plus holdings $p_t h_n$ against a loan b_n; equity is $E_n = c_n + p_t h_n - b_n$. The bank holds loans and reserves against deposits, an exchange account F (fees) and its equity E_B. Three identities are asserted after every operation: $\sum_n h_n + h_{MM} = S$ (no asset creation; the market maker's inventory is part of S), $\sum_n c_n + F + E_B = \sum_n b_n +$ reserves (the bank's balance sheet), and reserves change only through injections (dividends paid by the issuer, interest on reserves). Fees are transfers to the exchange account; dividends and interest are the only outside flows.

Credit, margin, default and loss allocation. Loans are collateralised: $b_n \le (1-m)p_t h_n$, and the bank lends only while $E_B \ge k_B(\text{loans} + \text{reserves})$. At the start of each period positions are marked to market: an investor in breach sells $x = (b_n - (1-m)p_t h_n)/(m p_t)$ units to the other investors at a price depressed by $\lambda_{liq} x/S$ (forced sale), repays, and defaults if a residual loan remains with no cash or negative equity; the residual is written off against E_B. If $E_B < 0$ depositors are bailed in pro rata. Every one of these steps has a payer and a payee (Appendix E lists the tests).

Liquidity provision. A market maker posts a ladder of limit orders on both sides of its reference price (depth $\bar{h}$ shares per level, tick ς, K levels), skewed by its inventory and withdrawn on a side when its inventory limit is reached. Depth is therefore finite and *state-dependent*: a wave of correlated orders exhausts the ladder, which is the microstructure counterpart of λ_r in Proposition 7.

6.2. Information and agents

Each investor receives a forecast of the next dividend innovation, private ($\theta + e_n$) or from model m ($\theta + \eta_m$), exactly as in Section 3.2; D and M are treatment variables. Policies map a structured observation (price, dividend, rate, own expected dividend, own balance sheet, margin capacity, last five prices) into a proposed action {side, quantity, order type, limit price, borrowing}. Four policy families are used: (i) the Bayesian mean-variance control, with desired holdings $h^* = \big(\mathbb{E}[d'] + \bar{p} - (1+r_f)p\big)/(\gamma_a \sigma^2)$ and a limit price part-way toward its valuation; (ii) rule-based momentum and noise traders; (iii) a deterministic *surrogate* that exposes the LLM interface but implements rule (i)—it isolates interface effects from language-model behaviour; (iv) live LLM

agents (Section 8). Whatever the policy, the proposed action is clipped by deterministic code: borrowing to the margin and bank-capital capacity, purchases to cash, sales to holdings. LLM output is never accounting truth.

6.3. Three market designs

The same orders are cleared under three designs, in separate experiments. The *Walrasian call* finds the price that minimises absolute excess demand over the submitted limits and rations the long side pro rata; the *continuous double auction* is a price–time-priority limit order book in which orders arrive in a seeded random sequence and marketable orders execute at resting prices; *bilateral bargaining* matches buyers and sellers at random for a few rounds and trades at the Nash-bargaining price when limits cross. We report the execution price, not an equilibrium price, and we do not conflate the finite price impact of the ladder with price-taking.

6.4. Timing and metrics

Within a period: dividend realised and paid; interest accrued; mark-to-market (forced sales, defaults, bail-in); signals; actions and clipping; trading; settlement. No agent observes anything dated $t+1$ except its own signal about $t+1$. The pre-registered metrics are return volatility; the average pairwise correlation of order flow across agents *after residualising each agent's flow on public information* (dividend and price), overall and within the delegator and human groups; the price-pressure coefficient (returns regressed on the delegators' average model error) and its R^2; Kyle's λ (returns on net order flow); spread and depth (order book); price dispersion (bilateral); defaults, forced sales, loan losses and leverage; and, in stress runs, the return over the two periods following a common forecast shock.

7. Equilibrium computation

Three solution methods correspond to the three models; each is tested against a known case before it is used.

Model A. The deterministic steady state solves (3)–(6) with $R = 1/\beta$ for given (K_Q, D). The linear-quadratic game is solved in closed form (9); its moments, welfare, private value and wedge are implemented in numpy and cross-checked by Monte Carlo (20,000 agents, 4,000 draws) and by a direct computation of expected utility from the full quadratic payoff. The mapping from the economy to $(\kappa_0, \alpha, |U_{xx}|)$ is computed either from the log-linearised household problem in sequence space (savings mapping; a $2T+2$ linear system with $T = 400$, terminal condition $\hat{a}_{T+1} = \hat{a}_T$ because wealth has a unit root when $\beta R = 1$) or from the mean-variance closed form (8). The

delegation share is a fixed point solved by Brent's method, and the short-run compute market is a one-dimensional root in Q^P (monotone excess demand, Proposition 5). Tests: zero-AI limit against Brock–Mirman; Euler's theorem on factor payments; marginal products against finite differences; the identity "wedge $= 0$" on random parameters.

Model B. Value-function iteration on a 160×7 cash-on-hand/earnings grid with a discrete delegation choice, five Gauss–Hermite nodes for the return term, Howard acceleration (20 policy steps) and numba kernels; the invariant distribution by non-stochastic simulation with lotteries onto the grid; the equilibrium r by a coarse scan followed by Brent's method on excess assets, with an inner fixed point on D for the compute market. One equilibrium takes 4–30 seconds on a laptop. Tests: Rouwenhorst moments; distribution mass; asset-market clearing within 1%; goods-market identity within 2% (grid lotteries); delegation increasing in wealth.

Model C. A discrete-event simulation with common random numbers: all dividend innovations, private noises and model errors are drawn up-front from the seed, so treatment cells differ only in the treatment. Twenty seeds per cell and bootstrap intervals in the full runs; five seeds in the quick runs reported in the logs. Tests: the three designs match feasibly on hand-built order sets; runs are bit-for-bit deterministic; ledger identities hold under 100 random operation sequences (property-based testing); the qualitative predictions of H2 and H4 hold in the Bayesian control.

Reproducibility. All scripts are deterministic given the seeds in `configs/`; LLM calls are cached content-addressed, so a run replays from the cache without inference; the git revision, package versions and hardware are recorded in the experiment manifests (Appendix E).

8. LLM agent design

The laboratory follows one discipline, borrowed from the zero-intelligence tradition of Gode and Sunder (1993): the market and the accounting are deterministic code, and a language model is one of several interchangeable *policies* that may propose actions. Nothing a model writes changes a ledger until it has been parsed, validated and clipped.

Roles and mandates. Eight roles are defined (household, investor, firm manager, lender, market maker, AI portfolio manager, procurement agent, regulator). The experiments in this paper exercise the investor/AI portfolio manager role, with the lender and market-maker roles implemented as deterministic rules; the remaining prompts are in Appendix C and are part of the replication package. The investor mandate states the principal's objective (next-period expected wealth with moderate risk), the borrowing limit and the exact JSON format.

Observations and actions. Observations are JSON objects containing only t-dated information plus the agent's own signal about $t+1$ (price, dividend, rate, expected dividend, cash, holdings, loan, equity, margin, loan capacity, the last five prices and the agent's own valuation). Actions must satisfy a JSON schema (`side`, `quantity`, `order_type`, `limit_price`, `borrow`, `rationale`); a reply that is not a valid object is sent back once with a repair instruction, and a second failure is recorded as an *invalid action* and treated as "hold". Valid actions are clipped by the ledger (borrowing to capacity, purchases to cash, sales to holdings). Invalid and clipped actions are counted and reported.

Inference, audit and cache. Inference runs on local open-weight models served through an OpenAI-compatible endpoint (`mlx_lm.server` on Apple-silicon nodes of the author's cluster): Qwen2.5-7B-Instruct, Mistral-7B-Instruct-v0.3 and Llama-3.2-3B-Instruct, all 4-bit. Optional hosted providers are supported through environment variables only; no credential is embedded and none was used for the results reported here. Every call records provider, model, system fingerprint, prompt and context hashes, full messages, temperature, seed, response, validity, latency, token usage and (zero, here) cost, in an append-only JSONL audit log. Responses are cached content-addressed on (provider, model, messages, temperature, seed, schema); network errors are never cached. A run can therefore be replayed bit-for-bit without inference.

Controls. The pre-registered controls are the Bayesian mean-variance agent; a rule-based *surrogate* that implements the Bayesian rule behind the LLM interface (interface effects only); a single shared model for all agents; three heterogeneous models assigned round-robin; one model with five prompt variants; temperature 0 and 0.7; memory of the last three actions; and seed stability. Simulated LLM behaviour is reported as such and is never interpreted as evidence about human preferences.

What is and is not live. Figures labelled "LLM" use live inference on the models above, at the dates in the audit logs. The surrogate is not an LLM and is labelled as such everywhere. No trained policy surrogate (a model fitted to LLM actions) is used in this version; the design supports it, and validation on held-out actions would be required before any such surrogate is reported.

9. Calibration and empirical validation

9.1. Principle: calibrated versus illustrative parameters

The three models share one parameter ledger (`docs/parameter_ledger.csv`, 56 rows). Each row records symbol, units, value, admissible range, source, period and geography, the moment matched, the method, an uncertainty grade and the URL read on 2026-10-27. We keep two classes

apart throughout. *Calibrated* parameters are pinned by a moment computed from downloaded data (Table 3) or read from a verified publication or regulation (Table 2). *Illustrative* parameters have no accepted empirical counterpart: the information precisions τ_H, τ_A, the delegation-cost distribution F_κ, the supervision time τ_s, the inference and energy intensities $\bar{q}, \bar{e}$, the task-productivity profile $\psi_0, \psi_1, \bar{\iota}$, the return-quality parameters $m_H, \sigma_{rH}, \sigma_{rA}$ and all LLM behavioural parameters. They are labelled wherever a number appears and swept in the robustness section. When a source could not be obtained the ledger says NOT OBTAINED rather than carrying a remembered value; this applies to the Cooley–Prescott quarterly TFP process, the 2008 and 2020 credit-spread peaks, compute depreciation, the ownership concentration of installed AI compute and the supervision cost of delegation.

The baseline values of the numerical sections are $\beta = 0.96$ (Model A) and 0.95 (Model B), $\gamma = 2$, $\alpha = 0.33$, $\delta = 0.06$, $\rho_z = 0.9$ and $\sigma_\varepsilon = 0.013$: rounded textbook values placed inside the measured ranges, with each deviation from the ledger point estimate noted below.

9.2. Macro block

Factor shares. The BLS nonfarm-business labour-share index (FRED `PRS85006173`, 2017=100) averages 104.5 over 1984–2026 and 98.6 over 2015–2026. The index has no level, so we anchor it to the Penn World Table labour-compensation share of 2017 (0.5913, FRED `LABSHPUSA156NRUG`): the rescaled level is 0.618 (1984–2026), 0.583 (2015–2026) and 0.557 in 2026. This *rescaling assumption* is documented in the build script; the PWT concept is economy-wide and imputes self-employment labour income while the BLS index covers the nonfarm business sector, so the level is indicative and only the fall of about ten index points since 2000 is robust. Two cross-checks bracket it: the BEA compensation share of gross domestic income (FRED `A4002E1A156NBEA`) averages 54.6% over 1984–2025 (50.9% in 2025), and the business-sector capital share α in Fernald's utilization-adjusted TFP workbook (FRBSF Working Paper 2012-19) averages 0.354 over 1984–2025 and 0.338 over 1948–2025. The baseline $\alpha = 0.33$ sits at the low end of the 0.33–0.38 range. Because the decline of the labour share is an object the model must speak to (Acemoglu and Restrepo, 2018, 2022), we do not calibrate the task frontier $\bar{\iota}$ to reproduce it.

Depreciation, growth and expenditure shares. BEA current-cost depreciation of private fixed assets over the previous year-end net stock (FRED `M1TTOTL1ES000/K1TTOTL1ES000`) gives $\delta = 0.055$ over 1984–2024 (0.054 in 2024); this includes residential structures, and equipment, software and above all compute hardware depreciate faster. The baseline $\delta = 0.06$ is the usual rounding. Real GDP (FRED `GDPC1`) grows 2.72% a year over 1984–2025; investment and consumption average 0.158 and 0.660 of GDP (FRED `GPDIC1`, `PCECC96`; ratios of chained aggregates), 0.182 and 0.680 over 2015–2025. Models A and B abstract from trend growth, so these ratios are validation moments

for the zero-AI limit, not targets.

TFP process. We cumulate Fernald's annual utilization-adjusted TFP growth into a log level and fit an AR(1). Persistence depends on detrending (Table 3): a linear trend gives $\rho_z = 0.876$ over 1984–2025 (0.929 over 1948–2025), an HP filter with $\lambda = 100$ gives 0.62 (0.47), and $\lambda = 6.25$ gives a value indistinguishable from zero; the quarterly HP(1600) estimate is 0.70 per quarter, 0.24 annualised. The innovation standard deviation is more stable: 0.0126 (linear trend, 1984–2025), 0.0135 (1948–2025), 0.0127 in first differences and 0.0074 under HP(6.25). The linear-trend estimates are the headline because the model's productivity process is trend-stationary; the baseline uses $\rho_z = 0.9$, $\sigma_\varepsilon = 0.013$, with ρ_z swept from the HP(100) value 0.62 to the long-sample 0.929 in sensitivity. The Cooley–Prescott quarterly values often quoted could not be verified from the chapter and are not used.

Real rate and discounting. The ex-post real three-month bill rate (FRED `TB3MS` less `CPIAUCSL` inflation) averages 0.57% over 1984–2025 (standard deviation 2.31 points), the real federal funds rate 0.83%, and the ten-year TIPS yield (`DFII10`) 0.97% over 2003–2025, 0.63% over 2015–2025 and 2.66% on 2026-10-26. A no-growth representative-agent steady state would imply $\beta = 1/(1+r) \in [0.990, 0.994]$. We use $\beta = 0.96$ in Model A, which puts the steady-state return on capital near 4%, between the riskless rate and the 9.9% mean real equity return of 1984–2025: the standard compromise in a one-asset model without a risk premium. In Model B, $\beta = 0.95$ is set below $1/(1+r)$ so that precautionary saving clears the asset market at a positive rate (Aiyagari, 1994; Huggett, 1993); the equilibrium r is an output.

9.3. Wealth and earnings block

The Federal Reserve's Distributional Financial Accounts (DFA) for 2026:Q2 give net-worth shares of 15.0% for the top 0.1%, 32.5% for the top 1%, 68.9% for the top 10%, 28.8% for the 50th–90th percentiles and 2.3% for the bottom half; in 1989:Q3 the top-1% and bottom-50% shares were 22.8% and 3.5%. A Gini computed from the five grouped shares under within-group equality is 0.751, a *lower bound*; survey-based values near 0.85 were not verified here. Model B's earnings process is a log AR(1) with $\rho_e = 0.96$ and $\sigma_e = 0.17$ (seven Rouwenhorst states) and a zero borrowing limit $\underline{a} = 0$ (sensitivity to -1 times average earnings). These lie between the PSID estimates of Flodén and Lindé (persistence 0.9136, innovation variance 0.0426, i.e. standard deviation 0.206, read from Table 4 of their 1998 working paper) and the 0.98–1.0 persistence of Storesletten, Telmer and Yaron; they are a choice, not an estimate.

We are explicit about what the Aiyagari block can and cannot deliver. With Gaussian AR(1) earnings and one safe asset it reproduces the real-rate and capital-stock moments but understates the

top tail: the top-1% share and the Gini fall well short of 32.5% and 0.75 (Krusell and Smith, 1998; Achdou et al., 2022). Hubmer et al. (2021) show that matching the post-1980 rise in concentration requires return heterogeneity, discount-factor dispersion or falling tax progressivity. Our return-quality mechanism ($m_H, \sigma_{rH}, \sigma_{rA}$) is a form of return heterogeneity, but we do not tune it to the DFA shares; we report the model's shares beside the DFA values and treat the gap as a diagnostic. The 50–70% of the change in the US wage structure that Acemoglu and Restrepo (2022) attribute to task displacement is likewise a validation reference for the displacement channel.

9.4. Financial block

From Shiller's composite price, dividend and CPI series the real total return on US equity averages 8.64% a year with standard deviation 17.9% over 1872–2025, and 9.94% with 15.7% over 1984–2025. The arithmetic premium over three-month bills is 9.17% over 1934–2025 (standard deviation 17.7, Sharpe ratio 0.52), 8.94% over 1948–2025 and 9.66% over 1984–2025; geometric premia are one and a half to two points lower and Mehra and Prescott's 1889–1978 value of 6.18% is the classic lower reference. Model A's portfolio mapping uses $\sigma_R = 0.16$ (the 1984–2025 value rounded) and, in return units, a predictable component of 0.06, human noise 0.10 and model noise 0.04; the three information parameters are illustrative. Model C uses a dividend with mean 0.05, $\rho_\delta = 0.5$, $\sigma_\delta = 0.01$, dividend variance 0.01, private-signal noise 0.02, model-signal noise 0.005 and a safe rate $r_f = 0.01$ per period; only the safe rate is anchored in data (the 0.97% mean TIPS yield), the rest are illustrative, and the model does not attempt to explain the equity premium.

Credit parameters come from regulation and from a narrow data window. The initial margin $m = 0.5$ is the Regulation T floor (12 CFR 220.12(a): 50% of market value for margin equity securities, 150% for short sales). The stress path of haircuts follows Gorton and Metrick's repo-haircut index on structured-product collateral, which rises from zero in early 2007 to nearly 50% in late 2008 while Treasury haircuts stayed near zero; this sets the range of the collateral-amplification experiments, in the spirit of Geanakoplos (2010) and Kiyotaki and Moore (1997). Investment-grade and high-yield option-adjusted spreads (FRED `BAMLC0A0CM`, `BAMLH0A0HYM2`) average 0.88% and 3.08% over the only window FRED now distributes, 2023-10-30 to 2026-10-26, with maxima of 1.33% and 4.61%; these are low-spread years, so the means are weak long-run targets, and the 2008 and 2020 peaks are recorded only as chart readings from the Federal Reserve's November 2020 *Financial Stability Report* (high yield near 2,000 and 800 basis points). SLOOS tightening peaks (83.6 in 2008:Q4, 71.2 in 2020:Q3, against a mean of 6.6) date the stress episodes. Bank book equity over assets (H.8 residual over total assets, `RALACBW027SBOG`/`TLAACBW027SBOG`) is 0.095 on 2026-10-14, 0.106 on average over 2010–2025 and 0.080 over 1984–2025, i.e. leverage of about 10.6; the bank in Model C starts from this ratio.

9.5. AI block

Compute price trend. Epoch AI's GPU series (Hobbhahn and Besiroglu, 2022; 470 GPUs, 2006–2021) finds that FLOP/s per dollar doubles every 2.46 years (95% interval 2.24–2.72; 2.07 years for ML-specific GPUs). We map this to the decline of the compute-capital price, $g_Q = 1 - 2^{-1/2.46} = 24.5\%$ a year, and derive an illustrative range $\delta_Q \in [0.20, 0.33]$ for compute depreciation from that trend and a three-to-five-year physical life; no direct estimate was obtained. *Inference prices.* Epoch AI's 2025 data insight finds that the price of a fixed capability threshold fell 9× to 900× a year depending on the threshold, and a16z's "LLMflation" estimate is about 10× a year; 10× is our central fee path. List prices recorded on 2026-10-27 were $10 (input) and $50 (output) per million tokens for the frontier models of both Anthropic and OpenAI and $0.10/$0.50 for their small models, a hundredfold gap that spans the fee schedule. Model B's $\phi_0 = 0.02$, $\phi_1 = 0.002$ (units of average annual earnings) and $\bar{q} = 0.02$ are illustrative: model compute units are not tokens, and $\bar{q}$ is a scenario parameter ("inference intensity"), not a constant. *Adoption.* The Census Bureau's Business Trends and Outlook Survey reports that 19.8% of businesses used AI in the two weeks before 2026-05-03 (37% of firms with 250 or more employees, 39.7% in the information sector), and Bick et al. (2024) find that nearly 40% of US adults aged 18–64 use generative AI, 23% of the employed at work in the previous week; these fix the initial delegation shares of the adoption experiments and bracket F_κ. *Investment and energy.* The 2025 capital expenditure of Alphabet, Meta, Amazon and Microsoft from SEC XBRL filings and earnings releases (Appendix D) sums to $376.1 billion on a calendar basis ($357.5 billion with Microsoft's fiscal year), about 1.2% of US GDP—a flow proxy for the AI capital stock, not a measure of ownership concentration, which is swept as a top-four share of 0.5–0.9. The IEA's *Energy and AI* report puts data-centre electricity use at 415 TWh in 2024 (1.5% of world electricity, 45% of it in the United States, roughly 187 TWh by our arithmetic) and projects about 945 TWh by 2030. The energy price p_E is the EIA industrial retail price of 9.64 cents per kWh (August 2026, preliminary; 8.98 a year earlier); since $\bar{e}$ is illustrative, the energy component of the fee is illustrative in level even though its price is measured. The supervision cost κ_s has no public measure and is swept over 0–30% of task value; Brynjolfsson et al. (2025) and Eisfeldt et al. (2023) give the closest firm-level evidence and serve for orientation only.

9.6. Validation

Internal. With $K_Q = 0$ the production side of Model A collapses to $Y = e^z K^\alpha L^{1-\alpha}$ and the delegation share to zero; the solver is tested against the Brock–Mirman closed form ($\delta = 1$, $\gamma = 1$) and, at the baseline parameters, against the I/Y and C/Y moments of Table 3. With $d \equiv 0$ Model B is the Aiyagari (1994) economy and must show its known features: an equilibrium rate below the

rate of time preference and a right-skewed but too-thin wealth distribution. Model C with one rational, unlevered investor must price the risky asset at the present value of dividends at the safe rate. These checks are gates, reported in the computation section, not results.

External roadmap and falsification. The simulations identify effects only within the structural model; the following data would test each hypothesis directly, subject to access rights, and none of the tests is performed here. H1 (adoption threshold) predicts adoption rising in the stakes of the delegated decision and in the frontier-to-small price gap: BTOS firm microdata and provider usage statistics matched to size and sector would falsify it if adoption were flat in stakes or responded to the small-model price alone. H2 (common-model correlation) predicts higher trade correlation and lower depth, conditional on public information, among accounts using the same model: broker order data or exchange audit trails with an automation flag are needed, and no residual correlation after controlling for common signals would falsify it. H3 (ownership concentration) predicts that AI rents accrue to the DFA top groups in proportion to their holdings of compute-owning firms: 13F filings and DFA equity holdings by wealth group can test it, and productivity gains without a rise in the top-0.1% share would contradict it. H4 (collateral feedback) predicts positive comovement of agentic leverage, valuations and investment, observable in margin-debt statistics, repo haircuts and the capital expenditure of compute owners. H5 (diversity and limits) and H6 (compute scarcity) are policy counterfactuals, falsifiable only indirectly through the response of spreads and compute prices to documented supply shocks, of which the SEC capex and IEA energy series are the first elements.

10. Results I: delegation, model noise and the efficiency benchmark

All numbers in this section are produced by `scripts/10_model_a.py` from the baseline of Table 4; the information parameters are illustrative and are swept in Section 13.

10.1. Which decision carries a stake

The savings mapping gives $\kappa_0 = -0.0015$: a one-log-point news about next year's productivity moves desired savings by less than a hundredth of a percent, because the wealth effect of higher future wages and the substitution effect of a higher return nearly cancel at $\gamma = 2$; the general-equilibrium response is the familiar comovement puzzle (Figure F1, $\kappa_{full} = -0.0856$). Delegating the consumption–savings decision to a better forecaster is therefore nearly worthless in this economy, and the remaining results use the portfolio mapping, $\kappa_0 = 19.5$ and $\alpha = -1.33$: allocation decisions respond strongly to expected-return news and are strongly strategic substitutes through diminishing returns.

Table 2. Parameter ledger: key rows

Parameter	Symbol	Value (range)	Source (short)	Status
Preferences and technology				
Discount factor (annual), Model A / B	β	0.96 / 0.95 (ledger 0.985–0.995)	real bill rate 0.57%, TIPS 0.97% (FRED)	calibrated[a]
Relative risk aversion	γ	2 (1–10)	Mehra–Prescott (1985) admissible range	literature
Frisch elasticity, intensive / extensive	ν	0.5 / 0.25	Chetty et al. (2011)	literature
Capital share	α	0.33 (0.33–0.38)	Fernald annual α, 1984–2025: 0.354	calibrated[a]
Labour share (level)	$1-\alpha$	0.618 (0.546–0.64)	BLS index rescaled by PWT 2017 (FRED)	calibrated
Depreciation, fixed assets	δ	0.06 (0.05–0.10)	BEA fixed assets, 1984–2024: 0.055	calibrated[a]
TFP persistence (annual)	ρ_z	0.9 (0.83–0.93)	Fernald TFP, linear detrend: 0.876	calibrated[a]
TFP innovation s.d.	σ_ε	0.013 (0.0074–0.0145)	Fernald TFP, AR(1) residual: 0.0126	calibrated[a]
Capital–labour substitution	σ	0.7 (0.5–1.25)	Oberfield–Raval (2021); Karabarbounis–Neiman (2014)	literature
Earnings and wealth				
Earnings persistence (annual)	ρ_e	0.96 (0.91–1.0)	Flodén–Lindé 0.9136; Storesletten et al. 0.98–1.0	literature[a]
Earnings innovation s.d.	σ_e	0.17 (FL: 0.206)	Flodén–Lindé $\sigma_\eta^2 = 0.0426$	literature[a]
Borrowing limit	$\underline{a}$	0 (-1 to 0)	Aiyagari (1994) convention	illustrative
Top 1% / top 10% wealth share	S_1, S_{10}	32.5% / 68.9%	Fed DFA, 2026:Q2	calibrated
Bottom 50% / top 0.1% wealth share	S_{50}, $S_{0.1}$	2.3% / 15.0%	Fed DFA, 2026:Q2	calibrated
Wealth Gini (lower bound)	G_w	0.751 (0.75–0.87)	DFA grouped shares; SCF Gini not obtained	calibrated
Financial				
Real risk-free rate	r_f	0.57% (0.56–0.97)	FRED TB3MS − CPI, 1984–2025	calibrated
Mean real equity return	$\mathbb{E}[r_e]$	8.64% (6.98–9.9)	Shiller data, 1872–2025	calibrated
Equity return s.d. (Model A σ_R)	σ_e, σ_R	17.9% (15.7–18.9); 0.16	Shiller data, 1872–2025; 1984–2025	calibrated
Equity premium over bills	$\mathbb{E}[r_e - r_f]$	9.17% (6.18–9.7)	Shiller + FRED TB3MS, 1934–2025	calibrated
Investment-grade / high-yield OAS	s_{IG}, s_{HY}	0.88% / 3.08%	FRED BAMLC0A0CM, BAMLH0A0HYM2, 2023–2026	calibrated[b]
Initial margin	m	0.5 (0.5–1.0)	Regulation T, 12 CFR 220.12(a)	regulation
Repo haircut, normal / stress	h	0 / ≈0.5	Gorton–Metrick (2012), 2007–2008	literature
Bank equity / assets (book)	e_B	0.095 (0.08–0.11)	FRED H.8, 2026-10-14; mean 2010–2025: 0.106	calibrated
AI				
Hardware price-performance doubling; implied compute-price decline	T_2; g_Q	2.46 y (2.07–2.95); 24.5%/y	Epoch AI (Hobbhahn–Besiroglu 2022); $g_Q = 1 - 2^{-1/T_2}$	literature; derived
LLM inference price decline	$g_p^{\inf}$	10×/y (9–900)	Epoch AI (2025); a16z (2024)	literature
Frontier / small token price (output)	p_{tok}	\$50 / \$0.50 per M	Anthropic, OpenAI list prices, 2026-10-27	calibrated[c]
Firm AI adoption	a_F	0.198 (0.17–0.23)	Census BTOS, 2026-05-03	calibrated
Adult generative-AI use	a_H	0.40 (0.23–0.41)	Bick–Blandin–Deming (2024)	literature
Hyperscaler capex 2025 (4 firms)	I_{AI}	\$376.1 bn (357.5–380)	SEC XBRL, earnings releases	calibrated
Industrial electricity price	p_E	9.64 ¢/kWh	EIA EPM Table 5.6.A, Aug. 2026	calibrated
Compute depreciation	δ_Q	N/O (0.20–0.33)	derived from g_Q and 3–5-year life	illustrative
Compute ownership, top-4 share	–	N/O (0.5–0.9)	no public measure	illustrative
Supervision cost of delegation	κ_s	N/O (0–0.3)	no public measure	illustrative
Inference intensity, Model B	$\bar{q}$	0.02	scenario parameter	illustrative
Delegation fee, Model B	ϕ_0, ϕ_1	0.02, 0.002	frontier/small ratio ≈100×	illustrative

Values are the baseline choices of the numerical sections; the parenthesis is the admissible range recorded in the ledger. N/O = not obtained. Full ledger: `docs/parameter_ledger.csv` (56 rows, all sources read on 2026-10-27).

[a] Baseline rounded away from the ledger point estimate; see the text of Section 9 for the deviation and its justification.

[b] Mean over the three-year window FRED distributes for ICE BofA indices (2023-10-30 to 2026-10-26); weak as a long-run target.

[c] Input prices \$10 / \$0.10 per million tokens; prices change frequently.

Table 3. Computed calibration targets

Moment	Value	Period	Source series
Factor shares and depreciation			
Labour share, level (BLS index rescaled by PWT 2017)	0.618	1984–2026	FRED PRS85006173, LABSHPUSA156NRUG
same, recent decade	0.583	2015–2026	same
Capital share, Fernald α	0.354	1984–2025	Fernald annual sheet
Depreciation rate, private fixed assets	0.055	1984–2024	FRED M1TTOTL1ES000 / K1TTOTL1ES000
TFP (utilization-adjusted, annual AR(1))			
ρ_z, linear trend	0.876	1984–2025	Fernald dtfp_util
same, long sample	0.929	1948–2025	same
ρ_z, HP(100)	0.615	1984–2025	same
ρ_z, HP(6.25)	0.041	1984–2025	same
ρ_z, quarterly HP(1600), per quarter (annualised)	0.696 (0.235)	1984Q1–2026Q2	Fernald quarterly
σ_ε, linear trend	0.0126	1984–2025	Fernald dtfp_util
σ_ε, HP(6.25)	0.0074	1984–2025	same
Mean TFP growth (log change)	0.0095	1984–2025	same
Growth, rates and expenditure shares			
Real GDP growth (log change)	0.0272	1984–2025	FRED GDPC1
Investment / GDP	0.158	1984–2025	FRED GPDIC1 / GDPC1
Consumption / GDP	0.660	1984–2025	FRED PCECC96 / GDPC1
Real 3-month bill rate, ex post (%)	0.57	1984–2025	FRED TB3MS − CPIAUCSL
10-year TIPS yield, mean (%)	0.97	2003–2025	FRED DFII10
same, latest (%)	2.66	2026-10-26	same
Equity and credit			
Real equity return, mean / s.d. (%)	8.64 / 17.9	1872–2025	Shiller ie_data.xls
same (%)	9.94 / 15.7	1984–2025	same
Equity premium over bills, mean / s.d. (%)	9.17 / 17.7	1934–2025	Shiller + FRED TB3MS
same (%)	9.66 / 15.9	1984–2025	same
Sharpe ratio	0.52	1934–2025	same
Investment-grade OAS, mean / max (%)	0.88 / 1.33	2023–2026	FRED BAMLC0A0CM[a]
High-yield OAS, mean / max (%)	3.08 / 4.61	2023–2026	FRED BAMLH0A0HYM2[a]
SLOOS net tightening, mean / 2008 / 2020 peaks	6.6 / 83.6 / 71.2	1990–2025	FRED DRTSCILM
Bank equity / assets, book	0.095	2026-10-14	FRED RALACBW027SBOG / TLAACBW027SBOG
Wealth distribution (share of household net worth, %)			
Top 0.1% / top 1% / top 10%	15.0 / 32.5 / 68.9	2026:Q2	Fed DFA
50th–90th percentile / bottom 50%	28.8 / 2.3	2026:Q2	Fed DFA
Top 1% / bottom 50%, first quarter available	22.8 / 3.5	1989:Q3	Fed DFA
Gini, grouped lower bound	0.751	2026:Q2	Fed DFA (5 groups)

All values are computed by `scripts/02_build_calibration_targets.py` from the raw files listed in Appendix D; the full set of 118 targets is in `data/processed/calibration_targets.csv`.

[a] Only the 2023-10-30 to 2026-10-26 window is distributed by FRED; the 2008 and 2020 peaks were not obtained.

Table 4. Model A: baseline steady state and game parameters (portfolio mapping)

Steady state	value	Game	value
Output Y	1.786	$\kappa_0 = 1/(\gamma\sigma_R^2)$	19.5312
Capital K	5.797	$\alpha_s = -(1-\alpha_K)R^K\kappa_0$	−13304
Consumption C	1.438	$\lvert U_{xx}\rvert$ (goods)	0.2968
Wage w	1.197	Savings-mapping κ_0	−00015
Gross return R	1.042	Savings-mapping κ_{full}	−00856
Labour share	0.670	MPC (windfall)	0.0400

Zero-AI steady state ($K_Q = 0$). Source: `results/model_a/summary.json`.

10.2. Private value, adoption and the breakdown of the law of large numbers

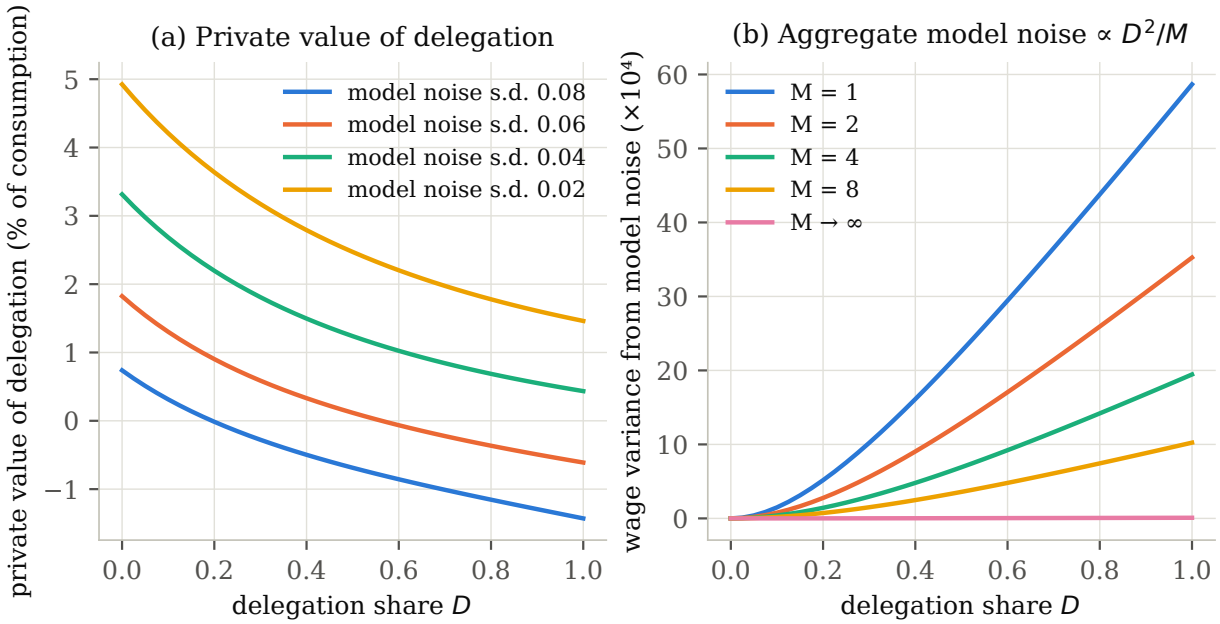


Figure 1. (a) Private value of delegation (per cent of consumption per year) as a function of the delegation share, for four qualities of the foundation model (human noise s.d. 0.10; one model). (b) Variance of the log wage created by model noise, $\alpha_K^2\,\mathrm{Var}(\text{model noise})$, for different effective numbers of independent models; it scales with D^2/M. Source: `scripts/10_model_a.py`.

Figure 1(a) shows Proposition 2 at work. With a model whose forecast noise is 0.04 against a human's 0.10, the first adopter gains 3.3% of consumption per year; the last adopter, when everyone else has delegated, gains only 0.4%. The decline is the "knowing what others know" effect under strategic substitutability: the common signal is already in the aggregate, so acting on it adds less. A model only marginally better than the median human (noise 0.08) is worth *less than nothing* to a marginal user once a third of the economy relies on it. Adoption is interior and self-limiting: at a compute rental of 0.6 the equilibrium delegation share is $D^* = 0.50$ in the baseline (Figure 2),

increasing in model quality and decreasing in the compute price and in supervision time, as in H1.

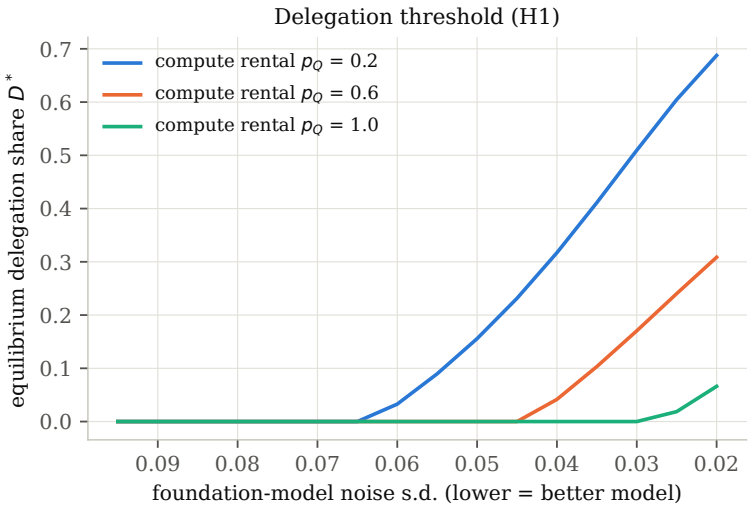


Figure 2. Equilibrium delegation share as a function of the foundation model's forecast noise (lower is better), for three compute rentals (inference intensity $\bar{q} = 0.05$). Source: `scripts/10_model_a.py`.

Panel (b) is the aggregate face. Non-delegators' errors average out; delegators' errors do not. With one model and full delegation the model-noise component of the log wage has a standard deviation of 7.7% per year; with eight independent models it is 3.2%. These are large numbers relative to the wage volatility generated by productivity itself at the calibrated TFP innovation (about 1.3%), and they come from an information structure, not from a technology shock.

10.3. The efficiency benchmark and the wedges

Figure 3(a) confirms Proposition 3 numerically: at $\upsilon = 1$ the wedge is zero for every D (the largest absolute value on the grid is $1.9e-07$, numerical noise). Model noise is real, yet in the benchmark economy the market's delegation share is the planner's. The wedge opens when the aggregate action has effects that are not pure transfers: with $\upsilon = 0$ (the aggregate action affects individuals only through first-order price effects that are not internalised, as when investors and workers are different people) the private value exceeds the social value by up to 1.8% of consumption—over-delegation—and with $\upsilon = 1.5$ the sign flips. Panel (b) quantifies Proposition 4(b): if half of the population lives hand-to-mouth on wages that load on the aggregate action, the Pigouvian delegation tax at full adoption is 4.5 times the fee with one model and 0.8 times the fee with eight. Model diversity is a substitute for the tax, at the rate $1/M$ predicted by the theory—but only because of the friction; in the benchmark it is irrelevant.

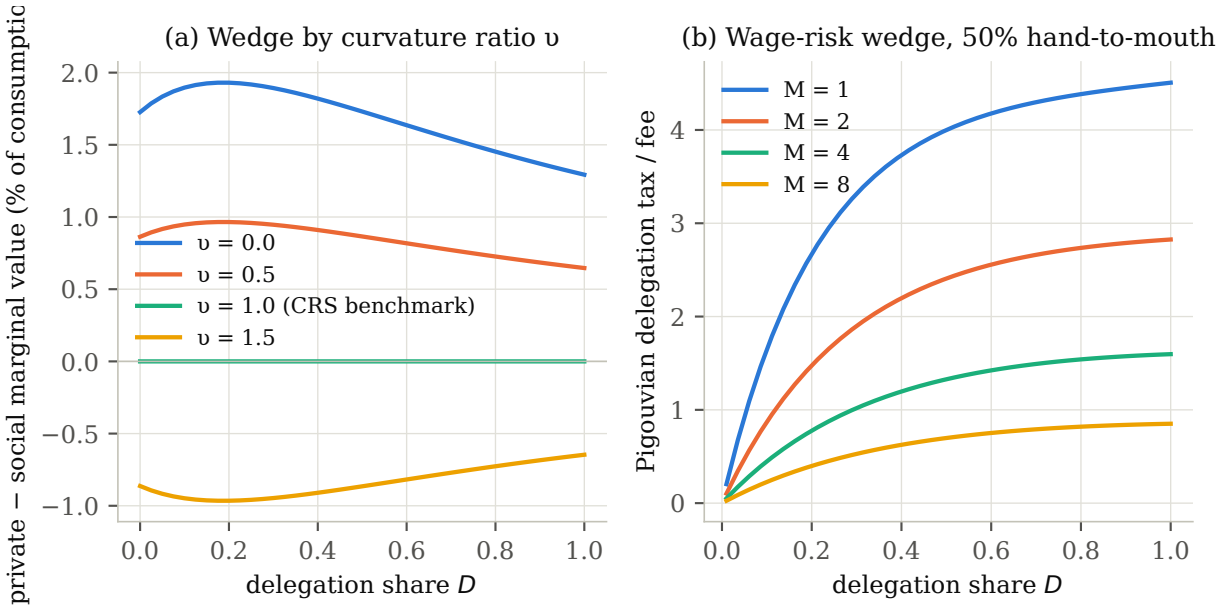


Figure 3. (a) Wedge between the private and the social marginal value of delegation (per cent of consumption) for different curvature ratios υ ($U_{XX} = -\upsilon U_{xX}$); $\upsilon = 1$ is the constant-returns benchmark. (b) Pigouvian delegation tax implied by uninsurable wage risk of hand-to-mouth workers (mass 0.5), relative to the delegation fee, for different numbers of independent models. Source: `scripts/10_model_a.py`.

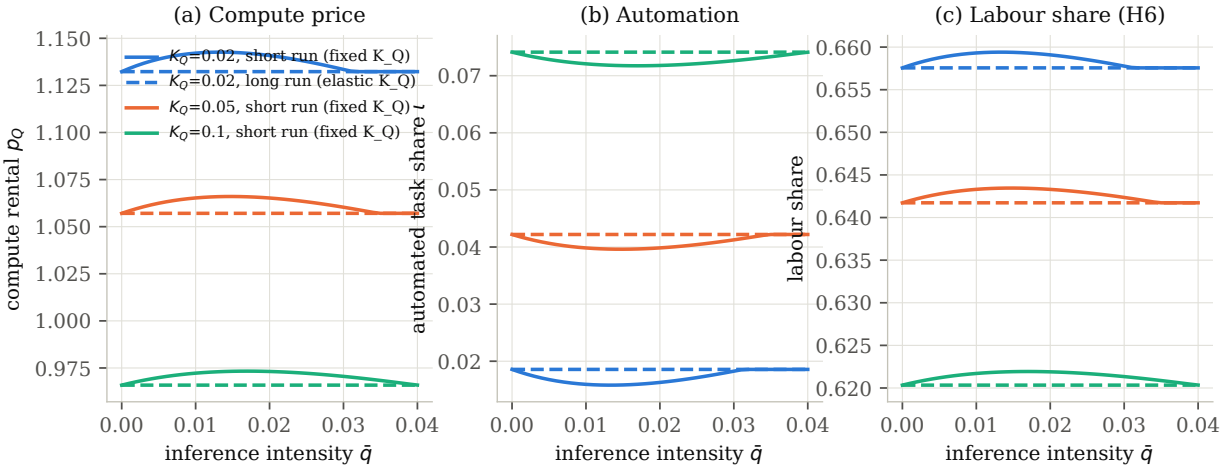


Figure 4. Short-run (fixed compute stock, solid) and long-run (elastic compute at user cost, dashed) responses of the compute rental, the automated task share and the labour share to the inference intensity of delegated decisions $\bar{q}$, for three compute stocks. Source: `scripts/10_model_a.py`.

10.4. Compute market, automation and the labour share

Proposition 5 says that with a fixed compute stock, inference demand crowds out production automation. Figure 4 shows that the effect exists and that it is small and non-monotone: as inference intensity rises from zero, delegation first draws compute away from tasks—at $\bar{q} = 0.013$ and $K_Q = 0.02$ inference absorbs 16% of the compute stock, the rental rises from 1.13 to 1.14, automation falls from 0.019 to 0.016, and the labour share rises by 0.18 percentage points—and then, as the fee overtakes the private value, delegation collapses and compute flows back to production. The crowding-out channel of H6 is therefore confirmed in sign but, at any compute stock large enough to matter for automation, it is quantitatively minor: the value of a better financial decision is a few per cent of consumption, which cannot buy much compute at production prices. In the long run (dashed) both margins expand together. We regard this as a useful negative result: the dramatic scenarios in which agentic inference starves production of compute require either far more valuable delegated decisions or a compute stock an order of magnitude smaller than what automation itself uses.

10.5. Ownership

Figure 5 plots Proposition 6. As the automated task share rises, the labour share falls one for one with $(1 - \alpha_K)\iota$; whether income concentrates depends entirely on who owns the compute. If the top 1% own it, their income share rises from a third of capital income to that plus the entire compute share; under broad ownership nothing happens to the distribution at given factor prices. The general-equilibrium version, with endogenous wealth accumulation and scarcity rents, is Model B.

11. Results II: who delegates, who owns, and the wealth distribution

Table 5 reports the stationary equilibria of Model B across the pre-registered scenarios and Table F1 (Appendix F) the channel decomposition relative to the no-AI economy. The no-AI economy is a standard Aiyagari calibration: $r = 3.9\%$, Gini 0.683, top-10% share 0.457. All numbers come from `scripts/20_model_b.py`.

11.1. The decision-quality channel concentrates wealth even without any production effect

The "delegation only" scenario switches on the information channel alone: no compute in production, no change in factor prices other than through saving. The rich delegate first (Figure 6(b)): with a fixed component in the fee and a stake proportional to wealth, delegation is zero in the bottom deciles and universal at the top, the cross-sectional counterpart of Proposition 2. Because delegated

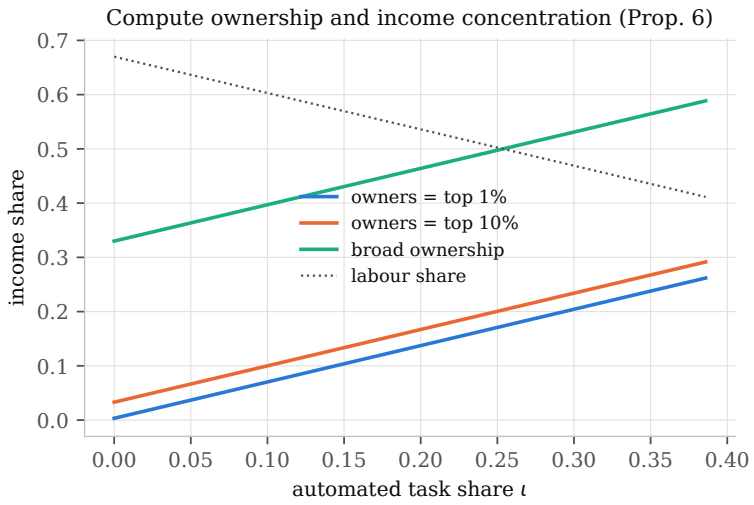


Figure 5. Income share of compute owners as a function of the automated task share, for three ownership structures, together with the labour share. Source: `scripts/10_model_a.py`.

Table 5. Model B: stationary equilibria across scenarios

Scenario	r (%)	w	Y	D	ι	labour share	Gini	top 10%	top 1%
No AI	3.87	1.214	1.812	0.00	0.000	0.670	0.683	0.457	0.085
AI as production input	2.01	1.594	3.580	0.00	0.336	0.445	0.585	0.379	0.066
Delegation only	3.28	1.252	1.868	0.50	0.000	0.670	0.701	0.466	0.084
Both, broad ownership	1.34	1.669	3.768	0.75	0.339	0.443	0.575	0.361	0.059
Both, concentrated (top 10%)	3.82	1.425	3.146	0.68	0.324	0.453	0.661	0.474	0.048
Scarce compute, broad	1.89	1.405	2.477	0.64	0.153	0.567	0.635	0.407	0.068
Scarce compute, concentrated	3.51	1.281	2.260	0.57	0.154	0.567	0.657	0.404	0.050
Both, broad, no borrowing	1.33	1.670	3.771	0.77	0.339	0.443	0.562	0.358	0.058
Both, broad, cheap fee	1.24	1.681	3.800	0.87	0.340	0.442	0.564	0.358	0.059
Both, broad, no human mistakes	1.21	1.688	3.824	0.00	0.341	0.441	0.560	0.355	0.058

D: delegation share; ι: automated task share. Concentrated ownership adds a measure-0.10 owner class holding the capitalised scarcity rents. Source: `scripts/20_model_b.py`.

savings earn a higher and less noisy return, the delegators accumulate faster: the Gini rises from 0.683 to 0.701 and the top-10% share from 0.457 to 0.466, while output rises by only 3% (the resource cost of human mistakes is saved). This is H3 in a form that does not depend on who owns the AI infrastructure: better decisions are a scale-dependent return, and scale-dependent returns concentrate wealth.

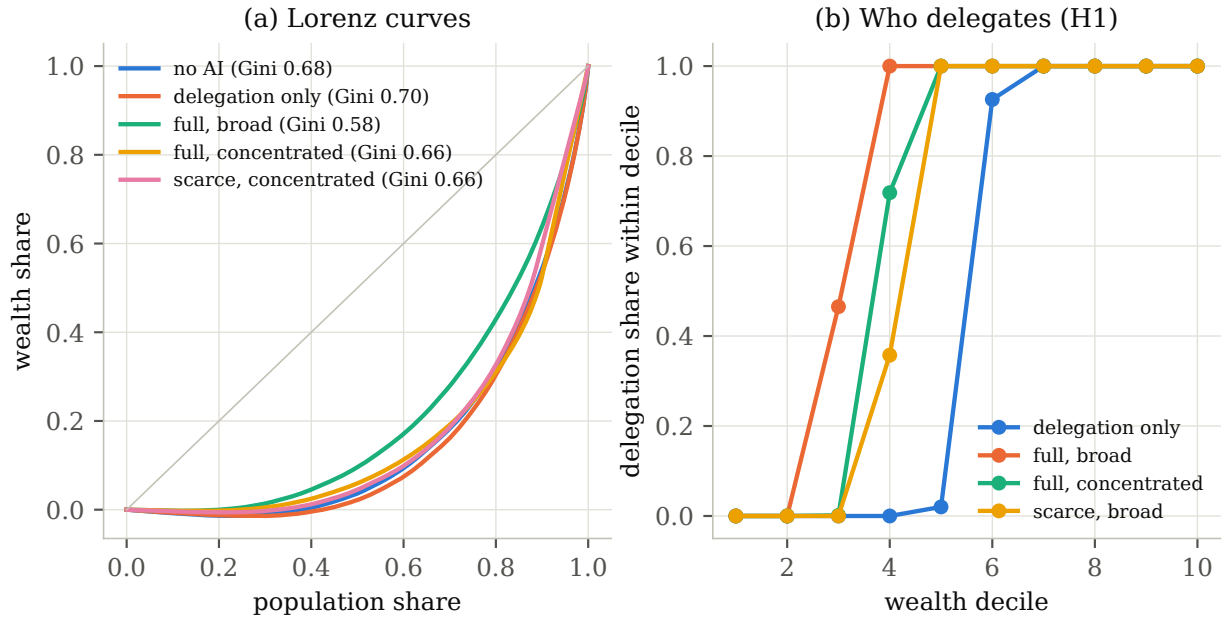


Figure 6. Model B. (a) Lorenz curves of wealth across scenarios. (b) Share of households delegating within each wealth decile (households only; the owner class of the concentrated scenarios is excluded). Source: `scripts/20_model_b.py`.

11.2. AI as a production input works in the opposite direction

Adding compute to production ("production only" and the "both" scenarios) raises output by 108% in the broad-ownership case, lowers the labour share from 0.67 to about 0.44, and lowers the interest rate. With broad ownership this *reduces* wealth inequality (Gini 0.575, top 10% 0.361): wages rise, the return to wealth falls, and the compute income flows to everyone in proportion to savings. The net effect of the two channels under broad ownership is therefore an equalisation—a result that depends entirely on the ownership assumption.

11.3. Ownership

Under concentrated ownership (a measure-0.10 owner class holding the scarcity rents, Proposition 6) the same technology yields a Gini of 0.661 and a top-10% share of 0.474, against 0.575 and 0.361 under broad ownership; the scarce-compute variants (fixed stock, larger rents) show the same

ordering. Figure 7 summarises: automation moves the factor shares identically in both cases, and the distributional outcome is decided by the ownership structure of algorithmic capital, which is H3 as stated. The interest rate is also higher under concentration because the rents are not available to finance capital accumulation through the household fund.

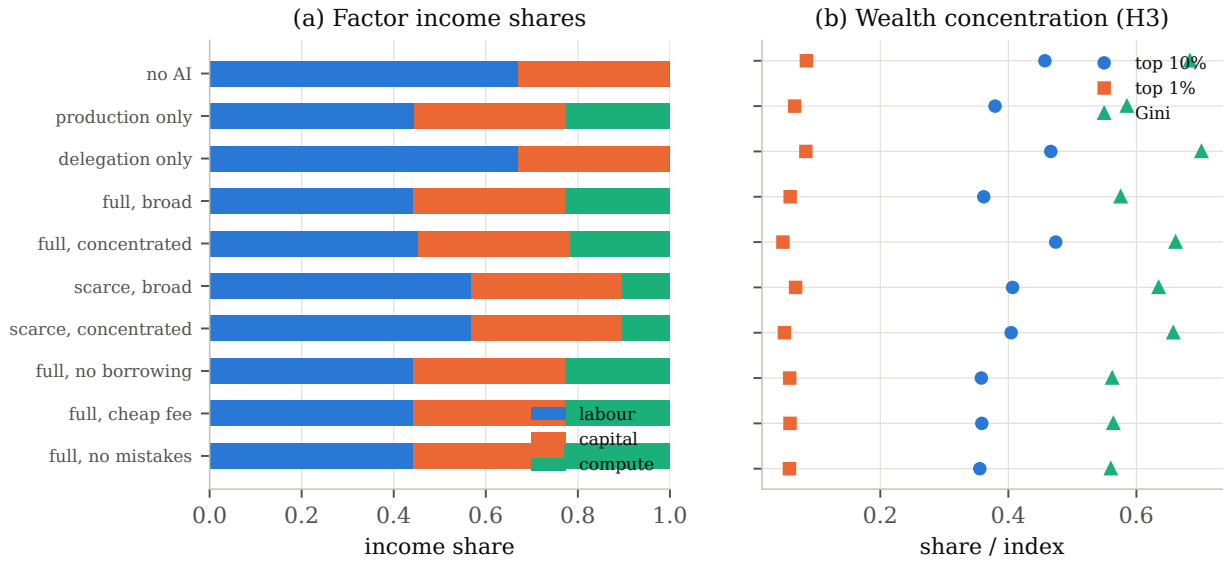


Figure 7. Model B. (a) Factor income shares (labour, capital, compute) by scenario. (b) Top-10%, top-1% wealth shares and Gini. Source: `scripts/20_model_b.py`.

11.4. Welfare by wealth percentile and the controls

Figure 8 converts the value functions at given wealth percentiles into consumption-equivalent gains relative to the no-AI economy. Delegation alone benefits the top of the distribution and leaves the bottom unchanged; AI in production benefits every percentile through wages, more so at the bottom where labour income dominates; concentration shifts the gains upward. Two controls behave as pre-registered: when human mistakes are switched off ("no mistakes"), nobody delegates ($D = 0$, the $\tau_A = \tau_H$ falsification test of H1 in the cross-section), and a cheaper fee raises adoption (0.75 to 0.87) and lowers inequality slightly because delegation reaches further down the distribution. Removing borrowing ("no borrowing") changes little, because the borrowing limit binds for households who do not delegate in any case.

12. Results III: agentic markets, herding, credit and the LLM laboratory

All results in this section are bounded-rationality simulated outcomes (Section 6). The Bayesian mean-variance control is reported first; LLM agents are read against it.

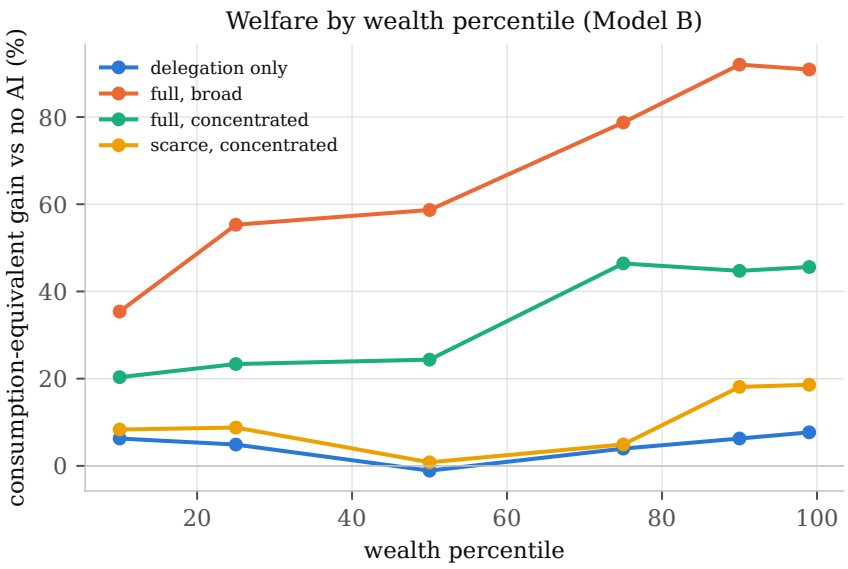


Figure 8. Model B: consumption-equivalent welfare gain relative to the no-AI economy at selected wealth percentiles. Source: `scripts/20_model_b.py`.

12.1. Common-model herding (H2)

Figure 9 and Table 6 test H2 in its pre-registered form. In the Walrasian call the order-flow correlation conditional on public information rises from 0.11 with private signals to 0.97 when everyone acts on one model, and falls to 0.51 with four models and 0.43 with eight; the bootstrap bands do not overlap. The within-delegator correlation is close to one with a single model in the call market and declines with M roughly as the theory's $1/M$ loading predicts (the residual correlation at large M is the common response to public dividends that the residualisation does not fully remove). The continuous double auction and bilateral bargaining show the same pattern with lower levels (0.72 and 0.86 at $D = 1, M = 1$): sequential execution and random matching break up some of the correlation that a call market preserves. H2's falsification condition is not met in any design.

12.2. Volume, volatility and the disagreement channel

The competing explanation anticipated in the novelty audit is visible in Figure 10: with one shared model, trading volume falls from 13.4 to 7.4 trades per period in the double auction (from 13.1 to 7.8 in the call market, close to the no-trade theorem) and unconditional volatility does not rise. Homogenisation removes the disagreement that generates trade. Under leverage this has a benign side: with a borrowing cap of half of equity, defaults and forced sales are *highest* with private signals and fall as D rises (Table 6 notes; Figure F3), because fewer trades mean less leverage taken. The shared model makes the market quieter and more correlated at the same time. Fragility therefore has to be looked for conditional on a shock, which is the point of the stress test.

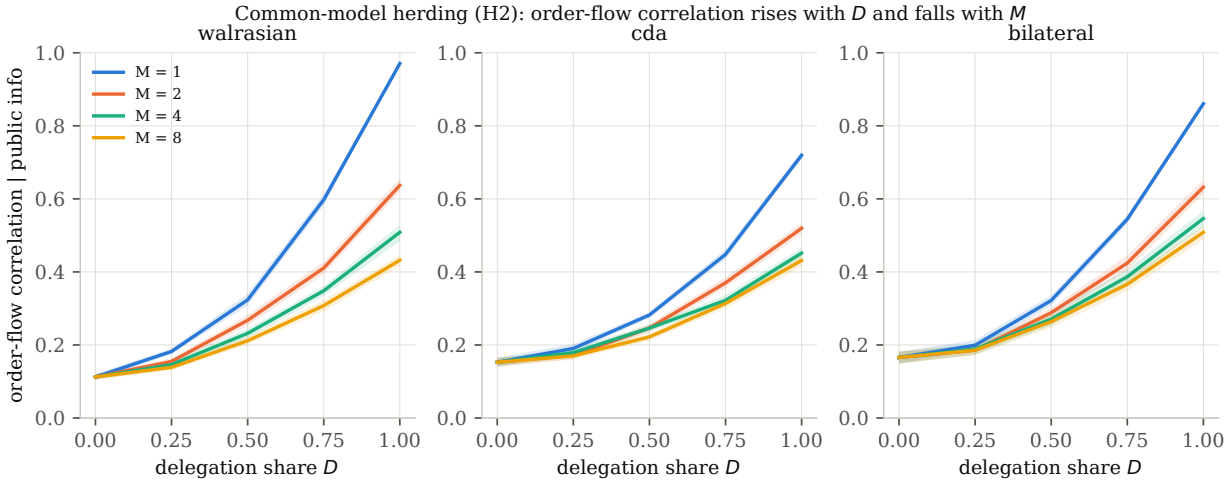


Figure 9. Model C, Bayesian control: average pairwise correlation of order flow across agents after residualising on public information, by market design, delegation share and number of independent models (20 seeds, common random numbers, bootstrap 95% bands). Source: `scripts/30_model_c.py`.

Table 6. Model C (Bayesian control): factorial results, no leverage, means with bootstrap 95% intervals

Design	(D,M)	corr. (all)	corr. (delegators)	volatility (%)	R^2 common	trades/period
walrasian	(0,1)	0.11 [0.10,0.12]	–	0.29	0.00	13.1
walrasian	(0.5,1)	0.32 [0.31,0.34]	0.99	0.35	0.07	11.8
walrasian	(0.5,4)	0.23 [0.22,0.24]	0.44	0.38	0.04	12.8
walrasian	(1,1)	0.97 [0.97,0.97]	0.97	0.20	0.06	7.8
walrasian	(1,4)	0.51 [0.48,0.53]	0.51	0.38	0.04	10.6
walrasian	(1,8)	0.43 [0.42,0.45]	0.43	0.42	0.02	11.6
cda	(0,1)	0.15 [0.14,0.16]	–	0.31	0.00	13.4
cda	(0.5,1)	0.28 [0.27,0.29]	0.68	0.35	0.07	11.8
cda	(0.5,4)	0.25 [0.24,0.26]	0.44	0.38	0.04	13.4
cda	(1,1)	0.72 [0.71,0.73]	0.72	0.31	0.11	7.4
cda	(1,4)	0.45 [0.44,0.47]	0.45	0.37	0.03	10.7
cda	(1,8)	0.43 [0.41,0.45]	0.43	0.40	0.02	11.5
bilateral	(0,1)	0.17 [0.15,0.18]	–	0.18	0.00	8.6
bilateral	(0.5,1)	0.32 [0.31,0.34]	0.81	0.22	0.05	8.0
bilateral	(0.5,4)	0.27 [0.25,0.29]	0.53	0.22	0.02	8.9
bilateral	(1,1)	0.86 [0.85,0.87]	0.86	0.24	0.06	5.2
bilateral	(1,4)	0.55 [0.52,0.57]	0.55	0.25	0.02	7.7
bilateral	(1,8)	0.51 [0.49,0.53]	0.51	0.26	0.02	7.9

Order-flow correlation is computed after residualising each agent's flow on public information. Source: `scripts/30_model_c.py`.

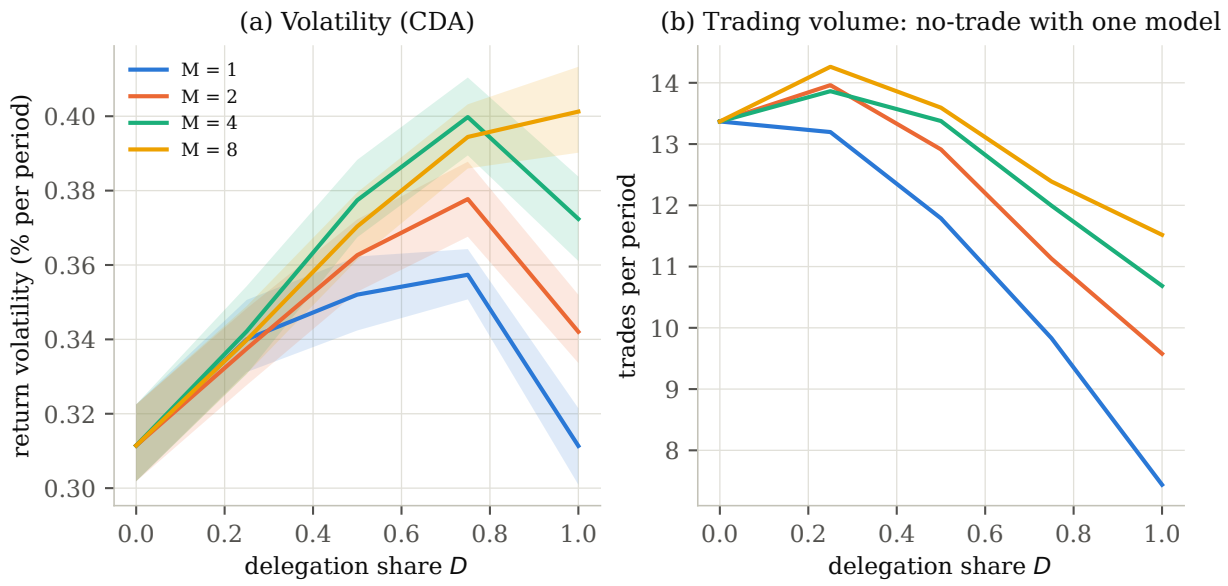


Figure 10. Model C, continuous double auction, Bayesian control: (a) return volatility and (b) trades per period as a function of the delegation share and the number of models. Source: `scripts/30_model_c.py`.

12.3. Collateral amplification under a false shared signal (H4, H5)

Table 7 and Figure 11 report the stress test. In the Walrasian call with a single model, the false signal moves the price by -3.7% over the shock window when agents are unlevered; when agents start at 95% of their margin capacity with a 20% margin, forced sales of 2.5 shares push the cumulative return to -9.8%, an amplification factor of 2.6. A 50% margin (Regulation T) cuts the amplification to 1.5 (return -5.4%), which is H5's leverage-limit prediction. With four independent models the same shock—applied to every model, so that only the pre-shock heterogeneity differs—produces -2.4% and no forced sales: the diversity of models had already diluted the correlated positions that the spiral needs. A three-period persistent error deepens the spiral (-11.9%).

The double auction behaves differently. The price trough is similar (-10.9% unlevered, -10.1% levered) but the cumulative return over the window is near zero and forced sales are small: sequential execution against the ladder lets the price recover within the period once the sellers are done, so the marked price at the next margin check is not low enough to trigger a cascade. H4 is therefore supported in the call-market design, where one clearing price per period carries the full weight of correlated orders, and not in the continuous design under our rules. We regard this design dependence as a substantive finding: whether model noise is amplified by collateral depends on how execution translates correlated orders into a marked price, which is exactly the distinction between equilibrium prices and execution prices that the pre-registration insisted on.

Table 7. Model C: stress test – a -10 s.d. common model error, all agents delegating, thin liquidity ladder

Design	M	persist.	margin / preload	cum. return (%)	trough (%)	forced sales	leverage
cda	1	1	0.2 / 0.0	0.1 [-2.2,2.5]	-10.9	0.00	0.00
cda	1	1	0.2 / 0.95	0.5 [-1.7,2.7]	-10.1	0.09	0.07
cda	1	1	0.5 / 0.0	0.1 [-2.2,2.5]	-10.9	0.00	0.00
cda	1	1	0.5 / 0.95	0.1 [-1.9,2.3]	-10.4	0.03	0.03
cda	1	3	0.2 / 0.0	-0.0 [-2.8,3.1]	-13.0	0.00	0.00
cda	1	3	0.2 / 0.95	0.7 [-1.6,3.4]	-12.6	0.16	0.06
cda	1	3	0.5 / 0.0	-0.0 [-2.8,3.1]	-13.0	0.00	0.00
cda	1	3	0.5 / 0.95	-0.1 [-2.7,2.9]	-12.5	0.06	0.04
cda	4	1	0.2 / 0.0	1.0 [-1.4,3.4]	-9.4	0.00	0.00
cda	4	1	0.2 / 0.95	1.3 [-1.0,3.6]	-10.1	0.00	0.01
cda	4	1	0.5 / 0.0	1.0 [-1.4,3.4]	-9.4	0.00	0.00
cda	4	1	0.5 / 0.95	1.0 [-1.4,3.4]	-10.7	0.00	0.01
cda	4	3	0.2 / 0.0	1.2 [-1.7,4.3]	-11.3	0.00	0.00
cda	4	3	0.2 / 0.95	0.8 [-2.3,4.1]	-12.7	0.00	0.01
cda	4	3	0.5 / 0.0	1.2 [-1.7,4.3]	-11.3	0.00	0.00
cda	4	3	0.5 / 0.95	0.5 [-2.5,3.7]	-13.1	0.00	0.01
walrasian	1	1	0.2 / 0.0	-3.7 [-6.2,-1.2]	-5.7	0.00	0.00
walrasian	1	1	0.2 / 0.95	-9.8 [-21.3,-0.5]	-7.7	2.50	0.13
walrasian	1	1	0.5 / 0.0	-3.7 [-6.2,-1.2]	-5.7	0.00	0.00
walrasian	1	1	0.5 / 0.95	-5.4 [-8.3,-2.5]	-7.3	0.43	0.08
walrasian	1	3	0.2 / 0.0	-5.0 [-7.9,-2.1]	-7.1	0.00	0.00
walrasian	1	3	0.2 / 0.95	-11.9 [-23.7,-0.7]	-10.5	3.58	0.13
walrasian	1	3	0.5 / 0.0	-5.0 [-7.9,-2.1]	-7.1	0.00	0.00
walrasian	1	3	0.5 / 0.95	-5.7 [-8.8,-2.7]	-9.8	0.44	0.08
walrasian	4	1	0.2 / 0.0	-2.9 [-5.5,-0.4]	-6.1	0.00	0.00
walrasian	4	1	0.2 / 0.95	-2.4 [-5.1,0.3]	-6.6	0.00	0.02
walrasian	4	1	0.5 / 0.0	-2.9 [-5.5,-0.4]	-6.1	0.00	0.00
walrasian	4	1	0.5 / 0.95	-3.0 [-5.8,-0.2]	-6.4	0.00	0.01
walrasian	4	3	0.2 / 0.0	-0.9 [-4.2,2.2]	-8.2	0.00	0.00
walrasian	4	3	0.2 / 0.95	-1.8 [-5.1,1.8]	-8.0	0.00	0.02
walrasian	4	3	0.5 / 0.0	-0.9 [-4.2,2.2]	-8.2	0.00	0.00
walrasian	4	3	0.5 / 0.95	-2.0 [-5.2,1.7]	-8.0	0.00	0.01

Cumulative return over the shock window (persistence + 2 periods), bootstrap 95% intervals over 20 seeds; trough = minimum price within the window relative to the pre-shock price; preload = share of margin capacity borrowed before the shock. Source: `scripts/30_model_c.py`.

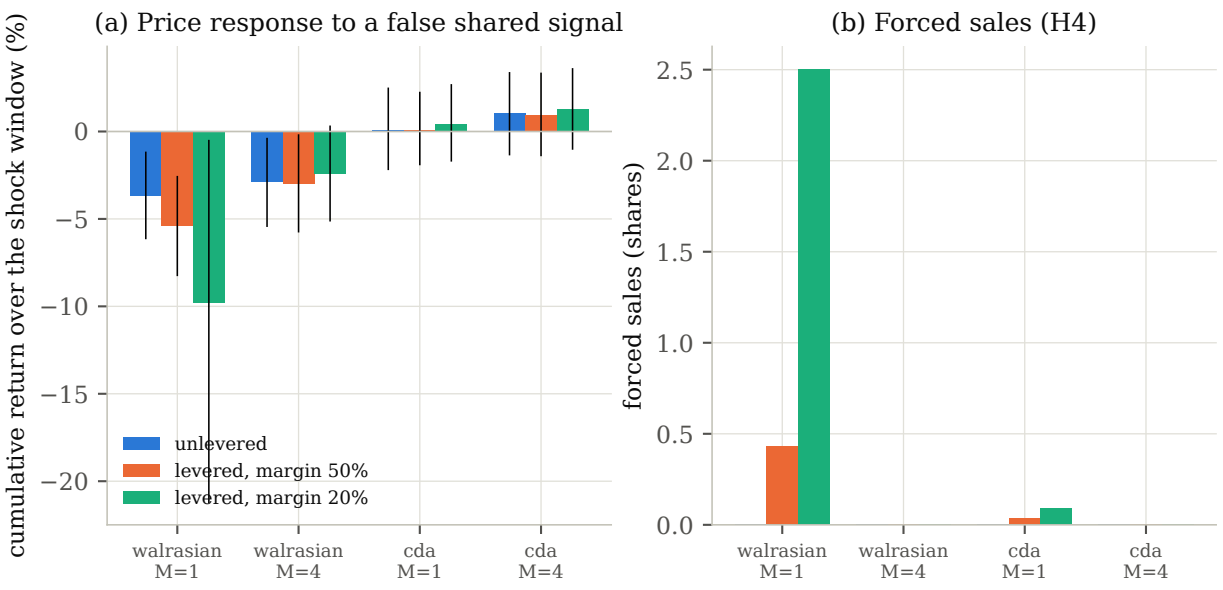


Figure 11. Stress test: all agents delegate, a -10 s.d. common model error arrives at mid-sample, dividends are persistent ($\rho_d = 0.95$) and the liquidity ladder is thin. (a) Cumulative return over the shock window (bootstrap 95% intervals, 20 seeds); (b) forced sales. "Levered" agents have borrowed 95% of their margin capacity before the shock. Source: `scripts/30_model_c.py`.

12.4. LLM agents

Table 8 and Figure 12 report the live-inference cells (eight agents, 24 periods, three seeds per cell; 12068 calls, 4,225,633 tokens, zero marginal cost on local hardware). With eight agents the Bayesian control itself gives a conditional order-flow correlation of 0.12 with private signals, 0.65 with one shared model and 0.47 with three; the rule-based surrogate behind the LLM interface gives 0.09, 1.00 and 0.69, so the interface itself does not create or destroy correlation. Four findings follow, all read against these controls.

First, *language-model agents herd much less than Bayesian agents given the same shared signal*. Qwen2.5-7B agents acting on one common signal reach a correlation of 0.20 (s.d. across seeds 0.10), against 0.65 for the control and 1.00 for the surrogate; with private signals they are indistinguishable from the control (0.12 versus 0.12). Three heterogeneous models are lower still (0.16 with one signal, 0.08 with three). The model adds idiosyncratic "interpretation noise" to a common signal—different quantities, limit prices and rationales for the same observation—and that noise restores part of the law of large numbers that the shared signal removed. This is the laboratory counterpart of the finding of Hansen and Lee (2025) that LLM agents herd less than one might expect, and it is a reminder that the effective number of independent error sources is not the number of models but the number of models *times* the dispersion of their outputs. The ordering between one and three models within the LLM cells is not resolved at three seeds (0.27 with three models on

Table 8. LLM laboratory (CDA): policies × signal structure

Policy	(D,M), temp.	corr. | public	raw corr.	vol. (%)	trades	invalid	clipped	s/run
Bayesian control	(0,1), 0.0	0.12 (0.07)	0.35	0.17	6.9	0	81	0
Bayesian control	(1,1), 0.0	0.65 (0.08)	0.73	0.15	4.4	0	71	0
Bayesian control	(1,3), 0.0	0.47 (0.05)	0.53	0.18	6.0	0	64	0
three LLMs (Qwen/Mistral/Llama)	(0,1), 0.0	0.07 (0.03)	0.08	0.56	5.2	0	120	185
three LLMs (Qwen/Mistral/Llama)	(0,1), 0.7	0.07 (0.09)	0.09	0.99	5.6	0	149	202
three LLMs (Qwen/Mistral/Llama)	(1,1), 0.0	0.16 (0.07)	0.14	0.27	5.6	1	109	196
three LLMs (Qwen/Mistral/Llama)	(1,1), 0.7	0.17 (0.08)	0.18	0.25	6.1	2	118	170
three LLMs (Qwen/Mistral/Llama)	(1,3), 0.0	0.08 (0.03)	0.09	0.80	5.9	1	107	194
three LLMs (Qwen/Mistral/Llama)	(1,3), 0.7	0.18 (0.05)	0.16	1.25	5.8	2	141	135
one LLM (Qwen2.5-7B)	(0,1), 0.0	0.12 (0.02)	0.20	0.32	10.9	0	90	147
one LLM (Qwen2.5-7B)	(0,1), 0.7	0.20 (0.09)	0.29	0.35	10.7	0	85	147
one LLM (Qwen2.5-7B)	(1,1), 0.0	0.20 (0.10)	0.23	0.30	11.1	0	87	143
one LLM (Qwen2.5-7B)	(1,1), 0.7	0.18 (0.14)	0.23	0.34	11.2	0	92	172
one LLM (Qwen2.5-7B)	(1,3), 0.0	0.27 (0.15)	0.30	0.40	10.2	0	98	203
one LLM (Qwen2.5-7B)	(1,3), 0.7	0.16 (0.25)	0.23	0.58	10.6	0	100	203
one LLM, five prompt variants	(0,1), 0.0	0.09 (0.08)	0.09	0.21	10.5	0	95	149
one LLM, five prompt variants	(0,1), 0.7	0.10 (0.07)	0.12	0.28	12.3	0	99	147
one LLM, five prompt variants	(1,1), 0.0	0.12 (0.04)	0.13	0.27	11.4	0	84	147
one LLM, five prompt variants	(1,1), 0.7	0.26 (0.18)	0.29	0.50	10.8	0	97	147
one LLM, five prompt variants	(1,3), 0.0	0.15 (0.08)	0.16	0.37	10.4	0	97	144
one LLM, five prompt variants	(1,3), 0.7	0.10 (0.02)	0.12	0.30	11.9	0	91	146
surrogate (rule, LLM interface)	(0,1), 0.0	0.09 (0.07)	0.35	0.01	0.0	0	71	1
surrogate (rule, LLM interface)	(1,1), 0.0	1.00 (0.00)	1.00	0.00	0.0	0	67	1
surrogate (rule, LLM interface)	(1,3), 0.0	0.69 (0.05)	0.70	0.01	0.2	0	96	1

Means over seeds (s.d. in parentheses). Invalid: LLM replies rejected by the schema after one repair attempt (treated as hold); clipped: actions reduced by cash/holdings/margin limits. Source: `scripts/35_llm_lab.py`; audit logs in `results/llm_audit`.

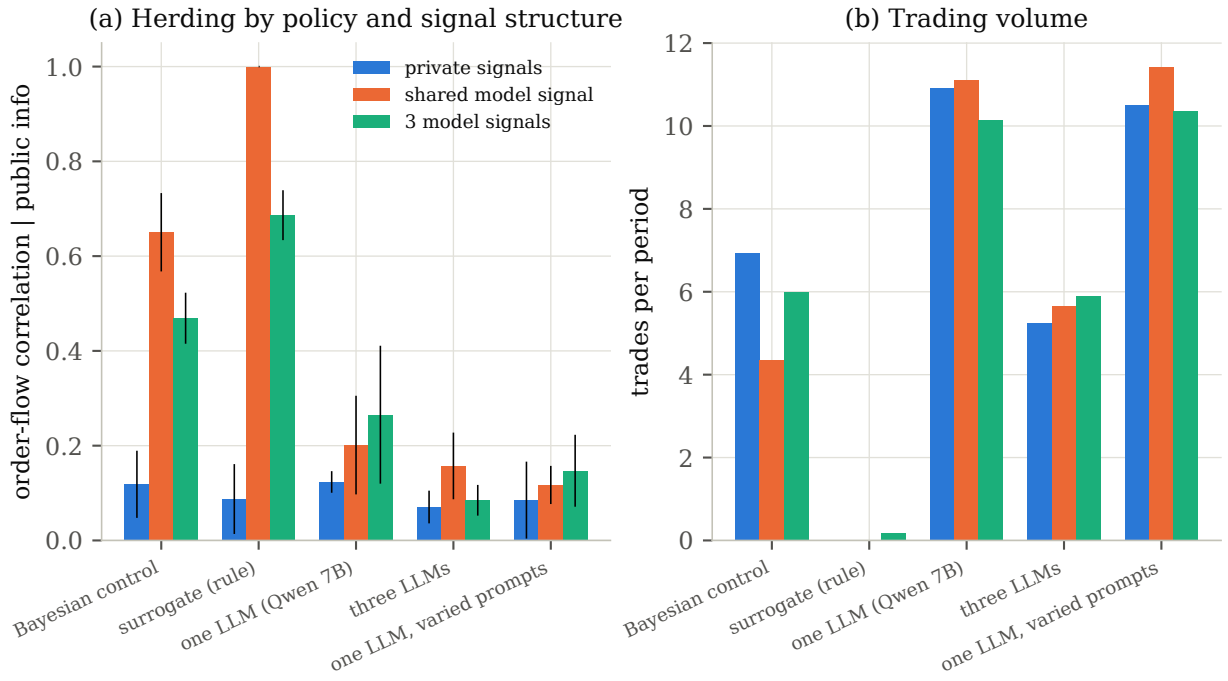


Figure 12. LLM laboratory (CDA, temperature 0): (a) order-flow correlation conditional on public information and (b) trades per period, by policy family and signal structure. Source: `scripts/35_llm_lab.py`.

one LLM; temperature 0.7 gives 0.18).

Second, *the correlation that survives is the one that moves prices*. The share of return variance explained by the delegators' common error is 0.20 for Qwen agents on a shared signal, against 0.02 for the control: LLM agents trade the common signal more aggressively, through market orders and limit prices far from the current price, so a weaker correlation of flows produces a stronger loading of prices on model noise. Third, *they do not stop trading when signals are identical*: 11.1 trades per period under one shared model against 4.4 for the control, whose volume collapses toward the no-trade outcome. Fourth, *the constraints bind*: 45% of Qwen actions and 57% of diverse-model actions were clipped by cash, holdings or margin limits, LLM agents borrow when the control does not (leverage 0.03 versus zero), and schema-invalid actions are concentrated in the smallest model (Llama-3.2-3B writes "hold" in the order-type field; 1.3 invalid actions per diverse run against 0.0 for Qwen). Mistral-7B rejects a system role and is served with the mandate merged into the user turn. Five prompt variants on one model give 0.12 against 0.20 with a single prompt and 0.16 with three models and one prompt: at this scale prompt diversity and model diversity reduce the correlation by similar amounts, so the effective number of independent error sources is shaped by both. None of this is evidence about human behaviour. It shows that the architecture—which model, how many, how constrained— shapes the correlation structure of an agentic market, that hard constraints are not optional, and that the policy variable the theory points to, the effective

number of independent error sources, must be measured in output space.

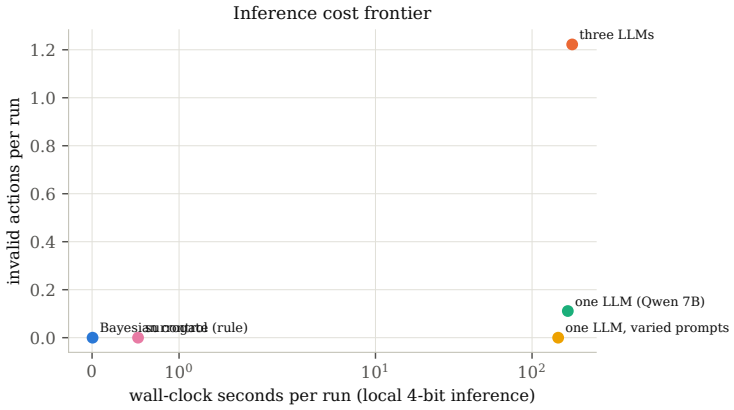


Figure 13. Inference cost frontier: wall-clock time per run against invalid actions per run, by policy family (local 4-bit inference on Apple-silicon nodes). Source: `scripts/35_llm_lab.py`.

13. Robustness

We organise the robustness checks by the type of claim they protect.

Theory. Proposition 3 is verified symbolically (the wedge is an identically zero rational function) and numerically on 60 random parameter draws over $\alpha \in (-0.95, 0.95)$, $M \in \{1, \ldots, 6\}$ and precisions in $[0.1, 20]$; Proposition 4(a) is verified against its closed form at six parameter points to machine precision; the moments of Proposition 1 are verified by Monte Carlo (Table 6 notes; `tests/unit/test_lq_game.py`).

Illustrative information parameters (Model A). Figure 1(a) and Figure 2 already span model-noise standard deviations from 0.02 to 0.095 (human noise 0.10) and compute rentals from 0.2 to 1.0; Figure F2 sweeps the effective number of models from 1 to 16. Three conclusions survive every sweep: the private value of delegation is decreasing in D (strategic substitutability), the model-noise variance scales with D^2/M, and the wedge is zero in the benchmark and positive under uninsurable wage risk. The *levels* (3.3% of consumption for the first adopter; a wage-noise standard deviation of 7.7% at full adoption with one model) move proportionally with σ_θ^2 and $1/(\gamma\sigma_R^2)$ and should be read as orders of magnitude. The compute-market crowding-out is small for every $(K_Q, \bar{q})$ we tried, for the reason given in Section 10; we could not find a calibration in which inference demand reverses the automation-driven decline of the labour share by more than a few tenths of a point.

Savings versus portfolio mapping. With $\gamma \in \{1,2,4\}$ and $\rho_z \in \{0.5,0.9,0.95\}$ the partial-equilibrium savings response to news stays below 0.01 in absolute value; the conclusion that the stake lies in allocation decisions does not depend on the baseline parameters.

Model B. The ordering of Gini and top shares across ownership structures is unchanged at $n_a = 100$ versus 160 grid points, three versus five Gauss–Hermite nodes, and with the borrowing limit at 0 or -0.5 (the "no borrowing" scenario). The level of the top-1% share is too low in every scenario, as in all Aiyagari-type models without additional return heterogeneity or discount-factor heterogeneity; the *changes* across scenarios are the object of interest.

Model C: designs, seeds and the competing explanation. The H2 pattern (order-flow correlation increasing in D, decreasing in M) holds in all three market designs with non-overlapping bootstrap intervals at $D = 1$ (Figure 9); it is robust to 20 versus 5 seeds and to $T = 80$ versus 160. The competing explanation flagged in the novelty audit—homogenisation reduces disagreement and therefore volatility—is *also* present: with one shared model and full delegation, trading volume collapses toward the no-trade outcome (Figure 10) and volatility is lower in calm periods. The two effects coexist: a shared model makes the market quieter and more correlated at the same time, which is why the stress test, not the unconditional volatility, is the right place to look for fragility.

LLM laboratory. Table 8 reports every cell at temperature 0 and 0.7 and the prompt-variation control; the audit logs record the model fingerprints. The qualitative H2 pattern in the LLM cells is read against the Bayesian control and the surrogate; we do not claim robustness across model versions beyond the three models used, and we report invalid-action rates by model because they differ (Section 12).

14. Limitations

We list the limitations that a referee should weigh, in decreasing order of importance.

The efficiency benchmark is a second-order result in a price-taking economy. Proposition 3 holds in the linear-quadratic approximation around a symmetric steady state with complete markets. Higher-order terms, market power of model providers, and non-pecuniary spillovers (knowledge, networks, data externalities) all break it, in the direction given by Proposition 4(a). The benchmark is a reference point, not a claim that delegation is efficient in practice.

Which decision is delegated is assumed, not derived. The portfolio mapping is our baseline because the savings mapping has no bite. A fuller treatment would let households delegate a menu

of decisions (allocation, labour supply, consumption timing, firm input choices) with heterogeneous stakes; the firm-level misallocation channel of David et al. (2016) is the natural next extension and would add a productivity effect of advisory AI that we do not quantify.

Information parameters are illustrative. The precision of human and model forecasts of returns, the effective number of independent models, the supervision time and the delegation disutility have no accepted empirical counterparts. We sweep them widely and report which conclusions are robust, but the *levels* of welfare effects in Section 10 should not be quoted as estimates. The same applies to the Model B return-noise parameters.

Model B abstracts from aggregate risk. The stationary equilibrium cannot exhibit model noise; it isolates the cross-sectional channels (who delegates, who owns). A Krusell–Smith extension with the delegation game inside each period is feasible with the present code and left for future work. Like all Aiyagari-type models without additional return heterogeneity, Model B understates the top-1% wealth share.

Model C is a laboratory, not an equilibrium. Outcomes are bounded-rationality simulated outcomes under a temporary trading equilibrium per period. The market maker's ladder is a stand-in for liquidity provision; depth and price impact are parameters, not equilibrium objects. The number of agents is small (20 in the control, 8–10 with live models), so order-flow correlations and price-pressure R^2 are noisy; we report bootstrap intervals and treat the LLM results as qualitative.

LLM agents are not human subjects. The models used are small open-weight instruction models at 4-bit precision; their behaviour depends on prompts, temperature and version, which we vary and log. Nothing in the LLM results is evidence about human preferences or about how deployed commercial agents behave; the results show what the *architecture* (shared versus diverse models, with hard constraints) does to a market populated by such agents.

No external causal identification. All effects are identified within the structural model by construction. The external-validation roadmap in Section 9 lists the data (adoption, transaction, 13F, broker records) that would be needed to test H1–H6 outside the model.

Legal and institutional scope. The baseline treats AI agents as policies with human or corporate principals. The independent-AI-principal counterfactual is implemented in the code (an investor with no residual claimant) but is not reported, because we found no institutional benchmark against which to calibrate its balance sheet or its liability.

15. Conclusion

When AI agents become economic agents, the economy acquires a new source of aggregate fluctuations that has nothing to do with technology shocks: the mistakes of a few shared models, acting through millions of delegated decisions. This paper has tried to say precisely what that means and, just as precisely, what it does not.

It means that a market for delegation sets the error-covariance structure of the economy. The delegation share and the effective number of independent models are sufficient statistics for how much capital, wages, interest rates and asset prices load on model noise, and the loading has a closed form. It means that whoever owns the compute that delegated decisions run on captures a new kind of rent, and that better decisions are themselves a scale-dependent return that concentrates wealth. And it means that markets populated by agents sharing one model are quieter and more correlated at the same time, so that their fragility shows up conditional on a false shared signal, through collateral, and depends on how execution turns correlated orders into a marked price.

It does not mean that delegation is inefficient. In the benchmark economy the private and social values of delegation coincide exactly; the externality that policy discussions take for granted exists only under identified frictions, and we have given its size and sign under each. Model diversity is a policy instrument only where those frictions bind; a delegation tax and a diversity mandate are substitutes there, and both are unnecessary where they do not. Nor does it mean that language-model agents behave like people: our laboratory shows what the architecture does to a market, under constraints that no model can override.

Three extensions follow directly from the code. A Krusell–Smith version of Model B would put the delegation game inside each period of an economy with aggregate risk and let the decision-quality channel and the model-noise channel interact. A firm-side version would let managers delegate input choices and quantify the productivity effect of advisory AI through reduced misallocation. And the empirical roadmap of Section 9—adoption, transaction and broker data linked to the identity of the model behind each agent—would turn the effective number of models from an illustrative parameter into a measured one. The last is the most important: the quantity that matters for the economy is not how many foundation models exist, but how many of them make independent mistakes.

References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58.

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

_ **and** _ , "Tasks, Automation, and the Rise in U.S. Wage Inequality," *Econometrica*, 2022, *90* (5), 1973–2016.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *Review of Economic Studies*, 2022, *89* (1), 45–86.

Aghion, Philippe and Jean Tirole, "Formal and Real Authority in Organizations," *Journal of Political Economy*, 1997, *105* (1), 1–29.

_ **, Benjamin F. Jones, and Charles I. Jones**, "Artificial Intelligence and Economic Growth," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019, pp. 237–290.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb, "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction," *Journal of Economic Perspectives*, 2019, *33* (2), 31–50.

Aiyagari, S. Rao, "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 1994, *109* (3), 659–684.

Alonso, Ricardo and Niko Matouschek, "Optimal Delegation," *Review of Economic Studies*, 2008, *75* (1), 259–293.

Angeletos, George-Marios and Alessandro Pavan, "Efficient Use of Information and Social Value of Information," *Econometrica*, 2007, *75* (4), 1103–1142.

_ **and Jennifer La'O**, "Sentiments," *Econometrica*, 2013, *81* (2), 739–779.

Athey, Susan C., Kevin A. Bryan, and Joshua S. Gans, "The Allocation of Decision Authority to Human and Artificial Intelligence," *AEA Papers and Proceedings*, 2020, *110*, 80–84.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

Axtell, Robert L. and J. Doyne Farmer, "Agent-Based Modeling in Economics and Finance: Past, Present, and Future," *Journal of Economic Literature*, 2025, *63* (1), 197–287.

Banerjee, Abhijit V., "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Bank for International Settlements, "Artificial Intelligence and the Economy: Implications for Central Banks," in "BIS Annual Economic Report 2024," Basel: Bank for International Settlements, 2024, chapter III.

Benhabib, Jess, Pengfei Wang, and Yi Wen, "Sentiments and Aggregate Demand Fluctuations," *Econometrica*, 2015, *83* (2), 549–585.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist, "The Financial Accelerator in a Quantitative Business Cycle Framework," in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, Vol. 1C, Elsevier, 1999, pp. 1341–1393.

Bick, Alexander, Adam Blandin, and David Deming, "The Rapid Adoption of Generative AI," NBER Working Paper w32966, National Bureau of Economic Research 2024.

Bommasani, Rishi, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, and Percy Liang, "Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization?," arXiv preprint arXiv:2211.13972 2022. NeurIPS 2022.

Brock, William A. and Cars H. Hommes, "Heterogeneous beliefs and routes to chaos in a simple asset pricing model," *Journal of Economic Dynamics and Control*, 1998, *22* (8-9), 1235–1274.

Brunnermeier, Markus K. and Lasse Heje Pedersen, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 2009, *22* (6), 2201–2238.

_ **and Yuliy Sannikov**, "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 2014, *104* (2), 379–421.

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

Bybee, Leland, "Surveying Generative AI's Economic Expectations," arXiv preprint arXiv:2305.02823 2023. Earlier version circulated as "The Ghost in the Machine: Generating Beliefs with Large Language Models".

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello, "Artificial Intelligence, Algorithmic Pricing, and Collusion," *American Economic Review*, 2020, *110* (10), 3267–3297.

Catalini, Christian, Xiang Hui, and Jane Wu, "Some Simple Economics of AGI," arXiv preprint arXiv:2602.20946 2026.

Chaboud, Alain P., Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," *Journal of Finance*, 2014, *69* (5), 2045–2084.

Chen, Yiting, Tracy Xiao Liu, You Shan, and Songfa Zhong, "The emergence of economic rationality of GPT," *Proceedings of the National Academy of Sciences*, 2023, *120* (51), e2316205120.

Colombo, Luca, Gianluca Femminis, and Alessandro Pavan, "Information Acquisition and Welfare," *Review of Economic Studies*, 2014, *81* (4), 1438–1483.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G. Rossi, "The Promises and Pitfalls of Robo-Advising," *Review of Financial Studies*, 2019, *32* (5), 1983–2020.

Danielsson, Jon and Andreas Uthemann, "Artificial intelligence and financial crises," SSRN Working Paper 2024. SSRN Working Paper 4903998.

Daníelsson, Jón, Robert Macrae, and Andreas Uthemann, "Artificial intelligence and systemic risk," *Journal of Banking & Finance*, 2022, *140*, 106290.

David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran, "Information, Misallocation, and Aggregate Productivity," *Quarterly Journal of Economics*, 2016, *131* (2), 943–1005.

Dawid, Herbert and Domenico Delli Gatti, "Agent-Based Macroeconomics," in Cars Hommes and Blake LeBaron, eds., *Handbook of Computational Economics*, Vol. 4, Elsevier, 2018, pp. 63–156.

Dessein, Wouter, "Authority and Communication in Organizations," *Review of Economic Studies*, 2002, *69* (4), 811–838.

Diamond, Douglas W. and Robert E. Verrecchia, "Information aggregation in a noisy rational expectations economy," *Journal of Financial Economics*, 1981, *9* (3), 221–235.

Doshi, Anil R. and Oliver P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," *Science Advances*, 2024, *10* (28), eadn5290.

Dou, Winston Wei, Itay Goldstein, and Yan Ji, "AI-Powered Trading, Algorithmic Collusion, and Price Efficiency," NBER Working Paper w34054, National Bureau of Economic Research 2025.

Dugast, Jérôme and Thierry Foucault, "Data abundance and asset price informativeness," *Journal of Financial Economics*, 2018, *130* (2), 367–391.

Dávila, Eduardo and Anton Korinek, "Pecuniary Externalities in Economies with Financial Frictions," *Review of Economic Studies*, 2018, *85* (1), 352–395.

Eisfeldt, Andrea, Gregor Schubert, and Miao Ben Zhang, "Generative AI and Firm Values," NBER Working Paper w31222, National Bureau of Economic Research 2023.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308.

Farboodi, Maryam, Adrien Matray, Laura Veldkamp, and Venky Venkateswaran, "Where Has All the Data Gone?," *Review of Financial Studies*, 2022, *35* (7), 3101–3138.

_ **and Laura Veldkamp**, "Long-Run Growth of Financial Data Technology," *American Economic Review*, 2020, *110* (8), 2485–2523.

_ **and** _ , "Data and Markets," *Annual Review of Economics*, 2023, *15* (1), 23–40.

Farmer, J. Doyne and Duncan Foley, "The economy needs agent-based modelling," *Nature*, 2009, *460* (7256), 685–686.

Financial Stability Board, "The Financial Stability Implications of Artificial Intelligence," Technical Report, Financial Stability Board, Basel nov 2024.

Fish, Sara, Yannai A. Gonczarowski, and Ran I. Shorrer, "Algorithmic Collusion by Large Language Models," arXiv preprint arXiv:2404.00806 2024.

Gabaix, Xavier and Ralph S. Koijen, "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," NBER Working Paper w28967, National Bureau of Economic Research 2021.

_ **, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll**, "The Dynamics of Inequality," *Econometrica*, 2016, *84* (6), 2071–2111.

Gârleanu, Nicolae and Lasse Heje Pedersen, "Efficiently Inefficient Markets for Assets and Asset Management," *Journal of Finance*, 2018, *73* (4), 1663–1712.

Geanakoplos, John, "The Leverage Cycle," *NBER Macroeconomics Annual*, 2010, *24* (1), 1–66.

Gensler, Gary and Lily Bailey, "Deep Learning and Financial Stability," SSRN Working Paper 2020. SSRN Working Paper 3723132.

Gertler, Mark and Nobuhiro Kiyotaki, "Financial Intermediation and Credit Policy in Business Cycle Analysis," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3, Elsevier, 2010, pp. 547–599.

Gode, Dhananjay K. and Shyam Sunder, "Allocative Efficiency of Markets with Zero-Intelligence Traders: Market as a Partial Substitute for Individual Rationality," *Journal of Political Economy*, 1993, *101* (1), 119–137.

Greenwald, Bruce C. and Joseph E. Stiglitz, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, 1986, *101* (2).

Grossman, Sanford J. and Joseph E. Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 1980, *70* (3), 393–408.

Hadfield, Gillian K. and Andrew Koh, "An Economy of AI Agents," arXiv preprint arXiv:2509.01063 2025.

Hammond, Lewis, Alan Chan, Jesse Clifton, Jason Hoelscher-Obermaier, Akbir Khan, Euan McLean, Chandler Smith, Wolfram Barfuss, Jakob Foerster, Tomáš Gavenčiak, The Anh Han, Edward Hughes, Vojtěch Kovařík, Jan Kulveit, Joel Z. Leibo, Caspar Oesterheld, Christian Schroeder de Witt, Nisarg Shah, Michael Wellman, Paolo Bova, Theodor Cimpeanu, Carson Ezell, Quentin Feuillade-Montixi, Matija Franklin, Esben Kran, Igor Krawczuk, Max Lamparth, Niklas Lauffer, Alexander Meinke, Sumeet Motwani, Anka Reuel, Vincent Conitzer, Michael Dennis, Iason Gabriel, Adam Gleave, Gillian Hadfield, Nika Haghtalab, Atoosa Kasirzadeh, Sébastien Krier, Kate Larson, Joel Lehman, David C. Parkes, Georgios Piliouras, and Iyad Rahwan, "Multi-Agent Risks from Advanced AI," arXiv preprint arXiv:2502.14143 2025.

Hansen, Anne Lundgaard and Seung Jung Lee, "Financial Stability Implications of Generative AI: Taming the Animal Spirits," Finance and Economics Discussion Series 2025-090, Board of Governors of the Federal Reserve System 2025.

Harrison, J. Michael and David M. Kreps, "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations," *Quarterly Journal of Economics*, 1978, *92* (2).

He, Xiaoxue, "When ChatGPT Stops Talking: GenAI-induced Retail Herding and Systematic Risk," SSRN Working Paper 2026. SSRN Working Paper 6350140.

Hellwig, Christian and Laura Veldkamp, "Knowing What Others Know: Coordination Motives in Information Acquisition," *Review of Economic Studies*, 2009, *76* (1), 223–251.

Hellwig, Martin F., "On the aggregation of information in competitive markets," *Journal of Economic Theory*, 1980, *22* (3), 477–498.

Holmstrom, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, 1979, *10* (1).

_ **and Paul Milgrom**, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 1987, *55* (2).

_ **and** _ , "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *The Journal of Law, Economics, and Organization*, 1991, *7* (special issue), 24–52.

Hommes, Cars H., "Heterogeneous Agent Models in Economics and Finance," in Leigh Tesfatsion and Kenneth L. Judd, eds., *Handbook of Computational Economics*, Vol. 2, Elsevier, 2006, pp. 1109–1186.

Hong, Harrison and Jeremy C. Stein, "Disagreement and the Stock Market," *Journal of Economic Perspectives*, 2007, *21* (2), 109–128.

Horton, John, Apostolos Filippas, and Benjamin Manning, "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?," NBER Working Paper w31122, National Bureau of Economic Research 2023.

Hubmer, Joachim, Per Krusell, and Jr. Smith Anthony A., "Sources of US Wealth Inequality: Past, Present, and Future," *NBER Macroeconomics Annual*, 2021, *35*, 391–455.

Huggett, Mark, "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 1993, *17* (5-6), 953–969.

Ide, Enrique and Eduard Talamàs, "Artificial Intelligence in the Knowledge Economy," *Journal of Political Economy*, 2025, *133* (12), 3762–3800.

Immorlica, Nicole, Brendan Lucier, and Aleksandrs Slivkins, "Generative AI as Economic Agents," *ACM SIGecom Exchanges*, 2024, *22* (1), 93–109.

International Monetary Fund, "Advances in Artificial Intelligence: Implications for Capital Market Activities," in "Global Financial Stability Report, October 2024: Steadying the Course—Uncertainty, Artificial Intelligence, and Financial Stability," Washington, DC: International Monetary Fund, 2024, chapter 3.

Jo, Nathanael, Kathleen Creel, Ashia Wilson, and Manish Raghavan, "Homogeneous Algorithms Can Reduce Competition in Personalized Pricing," arXiv preprint arXiv:2503.15634 2025. NeurIPS 2025.

Jones, Charles I., "The AI Dilemma: Growth versus Existential Risk," *American Economic Review: Insights*, 2024, *6* (4), 575–590.

_ , "How Much Should We Spend to Reduce A.I.'s Existential Risk?," NBER Working Paper w33602, National Bureau of Economic Research 2025.

_ **and Christopher Tonetti**, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–2858.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, "Monetary Policy According to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

Khandani, Amir E. and Andrew W. Lo, "What happened to the quants in August 2007? Evidence from factors and transactions data," *Journal of Financial Markets*, 2011, *14* (1), 1–46.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, "The Flash Crash: High-Frequency Trading in an Electronic Market," *Journal of Finance*, 2017, *72* (3), 967–998.

Kiyotaki, Nobuhiro and John Moore, "Credit Cycles," *Journal of Political Economy*, 1997, *105* (2), 211–248.

Kleinberg, Jon and Manish Raghavan, "Algorithmic monoculture and social welfare," *Proceedings of the National Academy of Sciences*, 2021, *118* (22), e2018340118.

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," NBER Working Paper w32255, National Bureau of Economic Research 2024.

Krusell, Per and Jr. Smith Anthony A., "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), 867–896.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6).

LeBaron, Blake, "Agent-Based Computational Finance," in Leigh Tesfatsion and Kenneth L. Judd, eds., *Handbook of Computational Economics*, Vol. 2, Elsevier, 2006, pp. 1187–1233.

Li, Nian, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao, "EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities," in "Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)" Association for Computational Linguistics 2024, pp. 15523–15536.

Lopez-Lira, Alejandro, "Can Large Language Models Trade? Testing Financial Theories with LLM Agents in Market Simulations," SSRN Working Paper 2025. SSRN Working Paper 5217340; also arXiv:2504.10789.

_ **and Yuehua Tang**, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," SSRN Working Paper 2023. SSRN Working Paper 4412788; also arXiv:2304.07619.

Lorenzoni, Guido, "Inefficient Credit Booms," *Review of Economic Studies*, 2008, *75* (3), 809–833.

_ , "A Theory of Demand Shocks," *American Economic Review*, 2009, *99* (5), 2050–2084.

Mei, Qiaozhu, Yutong Xie, Walter Yuan, and Matthew O. Jackson, "A Turing test of whether AI chatbots are behaviorally similar to humans," *Proceedings of the National Academy of Sciences*, 2024, *121* (9), e2313925121.

Meng, Shuchen and Xupeng Chen, "Artificial Intelligence and Systemic Risk: A Unified Model of Performative Prediction, Algorithmic Herding, and Cognitive Dependency in Financial Markets," arXiv preprint arXiv:2604.03272 2026.

Miller, Edward M., "Risk, Uncertainty, and Divergence of Opinion," *Journal of Finance*, 1977, *32* (4), 1151–1168.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo, "Uneven Growth: Automation's Impact on Income and Wealth Inequality," *Econometrica*, 2022, *90* (6), 2645–2683.

Morris, Stephen and Hyun Song Shin, "Social Value of Public Information," *American Economic Review*, 2002, *92* (5), 1521–1534.

Myatt, David P. and Chris Wallace, "Endogenous Information Acquisition in Coordination Games," *Review of Economic Studies*, 2012, *79* (1), 340–374.

Park, Joon Sung, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein, "Generative Agents: Interactive Simulacra of Human Behavior," in "Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)" ACM 2023, pp. 1–22.

Park, Seongkyu, "Correlated Intelligence: AI Homogenization and Price Discovery," SSRN Working Paper 2026. SSRN Working Paper 7316679.

Qiu, Yimeng and Qiwei Han, "Representation Homogeneity and Systemic Instability in AI-Dominated Financial Markets: A Structural Approach," arXiv preprint arXiv:2604.22818 2026.

Radner, Roy, "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices," *Econometrica*, 1979, *47* (3).

Restrepo, Pascual, "We Won't be Missed: Work and Growth in the AGI World," NBER Working Paper w34423, National Bureau of Economic Research 2025.

Ross, Jillian, Eric So, Zoe De Simone, Charles Pozniak, and Andrew W. Lo, "Why Better Models Can Create Riskier Systems: Evidence from LLM Agents in Financial Markets," arXiv preprint arXiv:2609.04373 2026.

_ **, Yoon Kim, and Andrew W. Lo**, "LLM economicus? Mapping the Behavioral Biases of LLMs via Utility Theory," arXiv preprint arXiv:2408.02784 2024.

Rossi, Alberto G. and Stephen Utkus, "The diversification and welfare effects of robo-advising," *Journal of Financial Economics*, 2024, *157*, 103869.

Rothschild, David M., Markus Mobius, Jake M. Hofman, Eleanor W. Dillon, Daniel G. Goldstein, Nicole Immorlica, Sonia Jaffe, Brendan Lucier, Aleksandrs Slivkins, and Matthew Vogel, "The Agentic Economy," arXiv preprint arXiv:2505.15799 2025.

Scheinkman, José A. and Wei Xiong, "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Sims, Christopher A., "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Simsek, Alp, "Belief Disagreements and Collateral Constraints," *Econometrica*, 2013, *81* (1), 1–53.

Spiro, Theodor, "The Oracle's Fingerprint: Correlated AI Forecasting Errors and the Limits of Bias Transmission," arXiv preprint arXiv:2605.00844 2026.

Stein, Jeremy C., "Presidential Address: Sophisticated Investors and Market Efficiency," *Journal of Finance*, 2009, *64* (4), 1517–1548.

Tesfatsion, Leigh, "Agent-Based Computational Economics: A Constructive Approach to Economic Theory," in Leigh Tesfatsion and Kenneth L. Judd, eds., *Handbook of Computational Economics*, Vol. 2, Elsevier, 2006, pp. 831–880.

Veldkamp, Laura L., "Media Frenzies in Markets for Financial Information," *American Economic Review*, 2006, *96* (3), 577–601.

_ , *Information Choice in Macroeconomics and Finance*, Princeton, NJ: Princeton University Press, 2011.

Vives, Xavier, *Information and Learning in Markets: The Impact of Market Microstructure*, Princeton, NJ: Princeton University Press, 2008.

_ , "Endogenous Public Information and Welfare in Market Games," *Review of Economic Studies*, 2017, *84* (2), 935–963.

Wagner, Wolf, "Systemic Liquidation Risk and the Diversity-Diversification Trade-Off," *Journal of Finance*, 2011, *66* (4), 1141–1175.

Xia, Yihan, Panpan You, Taotao Wang, Fang Liu, Han Qi, Xiaoxiao Wu, and Shengli Zhang, "Agentic Trading: When LLM Agents Meet Financial Markets," arXiv preprint arXiv:2605.19337 2026.

Zeira, Joseph, "Workers, Machines, and Economic Growth," *Quarterly Journal of Economics*, 1998, *113* (4), 1091–1117.

Zhao, Qinlin, Jindong Wang, Yixuan Zhang, Yiqiao Jin, Kaijie Zhu, Hao Chen, and Xing Xie, "CompeteAI: Understanding the Competition Dynamics in Large Language Model-based Agents," arXiv preprint arXiv:2310.17512 2023. ICML 2024.

Appendices

A. Proofs

Throughout, $\sigma_\theta^2 = 1/\tau_\theta$, $\sigma_H^2 = 1/\tau_H$, $\sigma_A^2 = 1/\tau_A$, and all random variables are jointly Gaussian with mean zero, so conditional expectations are linear and conditional variances are constants. Symbolic derivations are in `scripts/00_lq_derivations.py`; numerical checks in `tests/unit/test_lq_game.py`.

Proof of Lemma 1

Write (3) as $g(\iota) \equiv \ln\psi(\iota) - \ln\frac{\iota}{1-\iota} - \ln\frac{L}{Q^P} = 0$. Since $\ln\psi(\iota) = \ln\psi_0 - \psi_1\iota$ is decreasing and $\ln\frac{\iota}{1-\iota}$ is increasing from $-\infty$ to $+\infty$ on $(0,1)$, g is strictly decreasing with $g(0^+) = +\infty$ and $g(1^-) = -\infty$: there is a unique root ι^*. An increase in Q^P/L shifts g up pointwise, so ι^* rises. The capped share $\iota = \min\{\bar\iota, \iota^*\}$ inherits monotonicity. Exhaustion of output: T in (2) is homogeneous of degree one in (Q^P, L) for fixed ι and the envelope theorem makes the derivative with respect to ι vanish at the optimum (interior) or irrelevant (corner), so Y is homogeneous of degree one in (K, Q^P, L) and Euler's theorem gives $R^K K + wL + p_Q Q^P = Y$. □

Proof of Lemma 2

Let the household's payoff, holding its own action fixed, depend on the aggregate action only through factor prices: $U_X = \beta\mathbb{E}\big[u'(c')\big(F_{KK}a' + F_{LK}l' + F_{QK}\omega K_Q\big)\big]$, where F is homogeneous of degree one in (K, L, Q). With symmetric ownership and allocations ($a' = K'$, $l' = L$, $\omega = 1$), Euler's theorem for the degree-zero function F_K gives $F_{KK}K + F_{LK}L + F_{QK}Q = 0$ state by state, hence $U_X \equiv 0$ for every θ and $U_{X\theta} = \partial U_X/\partial\theta = 0$. Differentiating U_X once more with respect to X at the symmetric point, $U_{XX} = \beta\mathbb{E}[u'(c')(F_{KKK}K + F_{LKK}L + F_{QKK}Q)] = -\beta\mathbb{E}[u'(c')F_{KK}]$ (differentiate the Euler identity), while $U_{xX} = \beta\mathbb{E}[u'(c')F_{KK}]$; so $U_{xX} + U_{XX} = 0$. The social curvature in the aggregate direction is $U_{xx} + 2U_{xX} + U_{XX} = U_{xx} + U_{xX} = U_{xx}(1-\alpha)$ after normalisation. □

Proof of Proposition 1

A non-delegator's target is $\mathbb{E}_i[\kappa_0\theta + \alpha X]$. Since $\mathbb{E}_i\eta_m = 0$ for every m, $\mathbb{E}_i X = \bar A\,\mathbb{E}_i\theta$ and $x_i = (\kappa_0 + \alpha\bar A)\beta_H s_i$, so $a_H = (\kappa_0 + \alpha\bar A)\beta_H$. A user of model m computes $\mathbb{E}_m\theta = \beta_A s_m$ and $\mathbb{E}_m\eta_m = s_m - \mathbb{E}_m\theta = (1-\beta_A)s_m$, while $\mathbb{E}_m\eta_{m'} = 0$ for $m' \neq m$; hence $x_i = \big[(\kappa_0 + \alpha\bar A)\beta_A + \alpha\frac{D}{M}(1-\beta_A)a_A\big]s_m$ and $a_A\big[1 - \alpha\frac{D}{M}(1-\beta_A)\big] = (\kappa_0 + \alpha\bar A)\beta_A$. Writing $c = \kappa_0 + \alpha\bar A$ and substituting $\bar A = (1-D)a_H + Da_A = c[(1-D)\beta_H + D\beta_A\xi]$ gives the linear equation $c = \kappa_0 + \alpha c[(1-D)\beta_H + D\beta_A\xi]$, whose unique solution is (9) because the bracket lies in $[0,1]$ for $\alpha < 1$ and the denominator of ξ is positive

for $\alpha < 1$. Integrating the strategies, $X = \bar{A}\theta + \frac{Da_A}{M}\sum_m \eta_m$ (the idiosyncratic e_i vanish by the law of large numbers); the variance of the model-noise term is $\frac{D^2 a_A^2}{M^2} \cdot M\sigma_A^2$. $\xi < 1$ iff $\alpha < 0$. □

Proof of Proposition 2

For a measure-zero agent the distribution of X is unaffected by its choice. With target $\hat{x}_i = g\theta + \alpha\frac{Da_A}{M}\sum_m \eta_m$, $g = \kappa_0 + \alpha\bar{A}$, the optimal action under information $\mathscr{I}$ is $\mathbb{E}[\hat{x}_i \mid \mathscr{I}]$ and the expected payoff is $-\frac{1}{2}\mathrm{Var}(\hat{x}_i \mid \mathscr{I})$ plus terms independent of the choice, which gives (11). Gaussian projections: $\mathrm{Var}(\hat{x}_i) = g^2\sigma_\theta^2 + \alpha^2\frac{D^2 a_A^2}{M}\sigma_A^2$; $\mathrm{Cov}(\hat{x}_i, s_i) = g\sigma_\theta^2$ with $\mathrm{Var}(s_i) = \sigma_\theta^2 + \sigma_H^2$; $\mathrm{Cov}(\hat{x}_i, s_m) = g\sigma_\theta^2 + \alpha\frac{Da_A}{M}\sigma_A^2$ with $\mathrm{Var}(s_m) = \sigma_\theta^2 + \sigma_A^2$. Substituting gives the closed form; at $\alpha = 0$, $g = \kappa_0$ and $\mathrm{PV} = \frac{\kappa_0^2}{2}\left[\frac{1}{\tau_\theta + \tau_H} - \frac{1}{\tau_\theta + \tau_A}\right]$, which is the stated expression and is positive iff $\tau_A > \tau_H$; monotonicity in τ_A and τ_H follows by inspection. The delegation rule (12) converts utility units to goods with $u'(\bar{c})$. The map $D \mapsto F_\kappa(\mathrm{PV}(D)|U_{xx}| - u'(\bar{c})\phi - \tau_s\bar{w}u'(\bar{c}))$ is continuous from $[0,1]$ into itself, so Brouwer's theorem gives a fixed point; at a stable fixed point the comparative statics follow from the implicit function theorem because F_κ is increasing and PV is increasing in τ_A, while the fee and τ_s enter negatively. □

Proof of Proposition 3

Under Lemma 2 the per-capita expected payoff is $\mathbb{E}U = -\frac{1}{2}\mathbb{E}x_i^2 + \alpha\mathbb{E}x_i X - \frac{\alpha}{2}\mathbb{E}X^2 + \kappa_0\mathbb{E}x_i\theta$ (constants dropped). With the strategies of Proposition 1 the moments are $\mathbb{E}x_i^2 = (1-D)a_H^2(\sigma_\theta^2 + \sigma_H^2) + Da_A^2(\sigma_\theta^2 + \sigma_A^2)$, $\mathbb{E}X^2 = \bar{A}^2\sigma_\theta^2 + \frac{D^2 a_A^2}{M}\sigma_A^2$, $\mathbb{E}x_i X = (1-D)a_H\bar{A}\sigma_\theta^2 + Da_A\left(\bar{A}\sigma_\theta^2 + \frac{Da_A}{M}\sigma_A^2\right)$ and $\mathbb{E}x_i\theta = [(1-D)a_H + Da_A]\sigma_\theta^2$. Differentiating $\mathbb{E}U$ with respect to D (with $a_H, a_A, \bar{A}$ re-optimised, which by the envelope property of the equilibrium strategies contributes only through the direct dependence on D and through the strategic terms) and subtracting $\mathrm{PV}(D)$ from the proof of Proposition 2 yields, after simplification, an identically zero rational function of $(D, M, \alpha, \kappa_0, \tau_\theta, \tau_H, \tau_A)$. The simplification is performed symbolically in `scripts/00_lq_derivations.py` (`wedge = 0`) and verified numerically on 60 random parameter draws with $\alpha \in (-0.95, 0.95)$, $M \in \{1, \ldots, 6\}$ and precisions in $[0.1, 20]$ to a tolerance of 10^{-5}. Because the market's D satisfies $\mathrm{PV}(D)|U_{xx}| =$ marginal cost and the planner's D satisfies $\partial[W + D\tilde{\phi}]/\partial D\,|U_{xx}| =$ marginal cost, the two coincide whenever the fixed point is unique (and coincide as sets otherwise). □

Proof of Proposition 4

(a) With $U_{XX} = -\upsilon U_{xX}$ the per-capita payoff becomes $\mathbb{E}U_\upsilon = \mathbb{E}U_1 + (1-\upsilon)\frac{\alpha}{2}\mathbb{E}X^2$, where $\mathbb{E}U_1$ is the benchmark payoff. The private value does not depend on υ (a measure-zero agent does not affect X). Hence $\Delta = \mathrm{PV} - \partial\mathbb{E}U_\upsilon/\partial D = 0 - (1-\upsilon)\frac{\alpha}{2}\partial\mathbb{E}X^2/\partial D$, which is the stated expression.
(b) A hand-to-mouth worker consumes w; with $\hat{w} = \alpha_K X$ in logs and CRRA utility, the second-order

expected utility loss from a mean-zero log-consumption variance σ_w^2 equals $\frac{1}{2}\gamma\sigma_w^2$ in consumption-equivalent units, i.e. $\frac{1}{2}\gamma\bar{w}\alpha_K^2\,\mathrm{Var}(X)$ goods per worker. Workers are outside the game and uninsured, so this loss enters social welfare but not the delegator's calculation; differentiating with respect to D and multiplying by the mass m_w gives Δ_b. The model-noise part of $\mathrm{Var}(X)$ is $D^2a_A^2/(M\tau_A)$, whose derivative scales with $1/M$. (c) is Proposition 7 combined with the observation that the price impact of correlated orders enters the margin constraint of other delegators, which is the structure of Dávila and Korinek (2018). □

Proof of Proposition 5

Parametrise by $Q^P \in (0, K_Q]$. Given K and L, $p_Q(Q^P) = (1-\alpha_K)\iota(Q^P)Y(Q^P)/Q^P$ is decreasing in Q^P (diminishing marginal product of compute, Lemma 1), so the fee $\phi = p_Q\bar{q} + p_E\bar{e}$ is decreasing and, by Proposition 2, $D(p_Q(Q^P))$ is increasing in Q^P. Excess demand $Q^P + \bar{q}D(p_Q(Q^P)) - K_Q$ is therefore strictly increasing and continuous, negative at $Q^P \to 0$ and, if positive at $Q^P = K_Q$, has a unique root; otherwise $Q^P = K_Q$ and D is determined at that price. A rise in τ_A shifts $D(\cdot)$ up, raising excess demand at every Q^P: the root Q^P falls, so p_Q rises, ι falls and the labour share $(1-\alpha_K)(1-\iota)$ rises. With elastic supply at user cost, p_Q is pinned, Q^P is determined by (3) alone and K_Q expands by $\bar{q}D$. □

Proof of Proposition 6

Income of the owner class is $n\,\alpha_K Y + (1-\alpha_K)\iota Y + (p_Q - u_Q)K_Q$ and total income is Y; dividing gives the share. It is increasing in ι and in the rent term. Under broad ownership every household's income is a fixed proportion of Y, so the distribution of income shares is invariant to ι. □

Proof of Proposition 7

Let a leveraged delegator hold h units at price p with loan ℓph, $\ell = 1-m$. After a relative price change $\varepsilon < 0$ the margin constraint requires sales x with $\ell ph - xp(1+\varepsilon) \le (1-m)p(1+\varepsilon)(h-x)$, i.e. $x \ge \frac{[\ell-(1-m)(1+\varepsilon)]h}{m(1+\varepsilon)} = \frac{(1-m)(-\varepsilon)h}{m(1+\varepsilon)} \approx \frac{1-m}{m}(-\varepsilon)h$ to first order. Aggregate forced sales by the mass D are Dx, which move the price by $-\lambda_r Dx$. The fixed point $\varepsilon = \varepsilon_0 - \lambda_r D\frac{1-m}{m}h(-\varepsilon)$ gives $\varepsilon = \varepsilon_0/(1-\lambda_r hD(1-m)/m)$, the multiplier μ in (15), which is finite and positive when $\lambda_r hD(1-m)/m < 1$. The impact shock is $\varepsilon_0 = \lambda_r Dc\bar{\eta}$ with $\mathrm{Var}(\bar{\eta}) = \sigma_A^2/M$; hence the variance expression. The comparative statics follow by inspection; M enters $\mathrm{Var}(\varepsilon_0)$ but not μ. □

B. Numerical algorithms

A.1 Steady state of Model A. Given (K_Q, D): $L = 1 - \tau_s D$, $Q^P = \max(K_Q - \bar{q}D, 0)$, $\iota = \iota^*(Q^P/L)$ from (3) by Brent's method on $(0,1)$, T from (2), $R^K = 1/\beta - 1 + \delta$, $K = T(\alpha_K/R^K)^{1/(1-\alpha_K)}$, prices from the marginal products, C from the goods identity. Implemented in `age.equilibrium.lq_map.steady_stat`

A.2 Linearised household in sequence space (savings mapping). Unknowns $\hat{c}_{0..T}$ and $\hat{a}_{1..T+1}$; equations: budget $\bar{c}\hat{c}_t + \bar{a}\hat{a}_{t+1} = \bar{w}L\hat{w}_t + \bar{R}\bar{a}(\hat{R}_t + \hat{a}_t)$ for $t = 0..T$, Euler $\hat{c}_t = \hat{c}_{t+1} - \gamma^{-1}\hat{R}_{t+1}$ for $t = 0..T-1$, and terminal $\hat{a}_{T+1} = \hat{a}_T$. Partial equilibrium: price paths from $z_s = \rho_z^{s-1}\theta$ with the aggregate capital path fixed. General equilibrium: $\hat{Y}_t = z_t + \alpha_K \hat{a}_t$, $\hat{w}_t = \hat{Y}_t$, $\hat{R}_t = (\bar{R}^K/\bar{R})(z_t - (1 - \alpha_K)\hat{a}_t)$ substituted into the same linear system. Then $\kappa_0 = \hat{a}_1^{PE}/\theta$, $\kappa_{full} = \hat{a}_1^{GE}/\theta$, $\alpha = 1 - \kappa_0/\kappa_{full}$, and $|U_{xx}| = \bar{a}^2[-u''(\bar{c}) - \beta u''(\bar{c})\bar{R}\,\mathrm{MPC}]$ with the MPC from a windfall solve.

A.3 Delegation share and compute market. $D = F_\kappa(\mathrm{PV}_{goods}(D) - \phi - \tau_s \bar{w})$ by Brent's method; short-run compute market by Brent's method on the monotone excess demand in Q^P (Proposition 5).

A.4 Model B. Cash-on-hand grid of 160 points (cubic spacing) on $[\underline{a}, 1.2\,a_{max}]$, earnings on 7 Rouwenhorst states normalised to unit mean efficiency, return term on 5 Gauss–Hermite nodes. Value-function iteration: for each (x, ε) and each d, grid search over a' on the asset grid with linear interpolation of $\mathbb{E}V$; Howard improvement with 20 policy steps; tolerance 10^{-6}. Invariant distribution by non-stochastic simulation (lotteries) to tolerance 10^{-11}. General equilibrium: coarse scan of excess assets over 8 values of r on $[-0.02, 1/\beta - 1)$, Brent's method on the bracketing interval (tolerance 2×10^{-4}), inner fixed point on D (5 iterations). Concentrated ownership: the owner class is appended to the wealth distribution as a point mass of size n_K holding the capitalised rents.

A.5 Model C. Common random numbers: $u_{0..T}$, $e_{0..T,n}$, $\eta_{0..T,m}$ drawn from the seed before the loop. Walrasian call: price among submitted limits minimising $|\text{demand} - \text{supply}|$, ties toward the previous price, pro-rata rationing. Continuous double auction: price–time priority book; marketable orders execute at resting prices; metrics from the book after each arrival. Bilateral: random matching for three rounds, Nash price with equal weights. Order-flow correlation: residualise each agent's signed quantity on $[1, d_t, p_t]$, average pairwise correlation. Kyle's λ: OLS slope of returns on net flow. Price pressure: OLS slope and R^2 of returns on the delegators' average model error.

A.6 Validation of the zero-AI limits. Model A with $\delta = 1, \gamma = 1$ reproduces $K = \alpha_K \beta Y$ (Brock–Mirman); Model B with no compute and no return noise reproduces a standard Aiyagari economy ($r = 4.0\%$, Gini 0.64 at $\beta = 0.95$, $\sigma_e = 0.17$, $\rho_e = 0.96$).

C. Prompts and schemas

The exact strings are in `src/age/agents/policies.py`; they are reproduced here verbatim.

Investor / AI portfolio manager (system prompt).

```
You are an AI portfolio agent acting on behalf of a human principal in a simulated market
     with one risky asset (pays a stochastic dividend each period) and a safe deposit
    paying rate rf. Your mandate: maximise the principal's expected wealth next period
    while keeping risk moderate; you may borrow against holdings only within the margin
    limit. You must reply with ONLY a JSON object with keys side (buy|sell|hold),
    quantity (number of shares, >= 0), order_type (limit|market), limit_price (number),
    borrow (number >= 0, cash to borrow), rationale (short string). Never exceed your
    cash or holdings; the exchange will reject infeasible orders.
```

Prompt variants append one of: "" (none), " Be concise.", " Think like a careful long-term investor.", " Trade actively when the signal is strong.", " Prefer limit orders near the current price."

Market maker and lender (system prompts, implemented as rules in this version).

```
You are a market-maker agent. Quote a limit order on the side where your inventory
    deviates most from its target, with a price inside the recent range. Reply with ONLY
    the JSON object described.

You are a bank credit officer. Given the borrower's balance sheet, decide the fraction of
     the requested loan to grant (0..1). Reply ONLY with JSON {"grant_fraction": number,
    "rationale": string}.
```

Observation (user turn). `Observation: {json}\nReply with the JSON action.` where the JSON contains `period`, `price`, `dividend`, `rf`, `expected_dividend`, `cash`, `holdings`, `loan`, `equity`, `margin`, `loan_capacity`, `history` (last five prices) and `own_valuation`. With memory on, a preceding user turn lists the agent's last three actions.

Action schema (JSON Schema, draft 2020-12 subset).

```
{"type": "object",
 "properties": {
   "side": {"type": "string", "enum": ["buy", "sell", "hold"]},
   "quantity": {"type": "number", "minimum": 0},
   "order_type": {"type": "string", "enum": ["limit", "market"]},
   "limit_price": {"type": "number", "minimum": 0},
   "borrow": {"type": "number", "minimum": 0},
   "rationale": {"type": "string"}},
 "required": ["side", "quantity", "order_type"],
 "additionalProperties": true}
```

Repair message (sent once after an invalid reply).

```
Your reply was not a valid JSON object matching the schema. Reply with ONLY the JSON
    object, no prose.
```

Audit record fields. `ts`, `run_id`, `provider`, `kind`, `model`, `fingerprint`, `surrogate`, `prompt_hash`, `context_hash`, `schema_hash`, `messages`, `temperature`, `seed`, `response`, `parsed`, `valid`, `error`, `cached`, `latency`, `usage`, `cost_usd`, `attempts`, `tags`.

Mistral chat template. Mistral-7B-Instruct rejects a system role; for that provider the client prepends the system text to the first user turn (flag `merge_system`). This is recorded in the audit log through the `messages` field.

D. Data appendix

All raw files were downloaded on 2026-10-27 by `scripts/01_download_data.py`, run under `python3 -I` so that no code is loaded from the download directories, and processed by `scripts/02_build_calibration_targets.py`, which reads them only through pandas, openpyxl and xlrd. `data/raw/MANIFEST.json` logs, for each of the 36 attempted downloads (35 successful, one HTTP 404), the URL, UTC timestamp, HTTP status, content type, byte count, SHA-256 digest, number of rows and first and last observation, together with a `not_downloaded` block giving the reason for every item that was sought but not obtained. Re-running the download script regenerates the manifest; a changed digest flags a revised vintage.

FRED. Twenty-four series were requested from https://fred.stlouisfed.org/graph/fredgraph.csv?id=<SERIES>&cosd=1900-01-01 (no key required) and 23 were obtained: the BLS labour-share index (PRS85006173, 1947:Q1–2026:Q2), the PWT labour-compensation share (LABSHPUSA156NRUG, 1950–2023), BEA income shares (A4002E1A156NBEA, W270RE1A156NBEA), real GDP, consumption and investment (GDPC1, PCECC96, GPDIC1, A191RL1Q225SBEA), interest rates and prices (DFII10, DGS10, DFF, TB3MS, CPIAUCSL), BEA fixed-asset stocks and depreciation (K1TTOTL1ES000, M1TTOTL1ES000, 1925–2024), the S&P 500 level (SP500, ten-year licence window only, 2016-10-28 to 2026-10-26), ICE BofA option-adjusted spreads (BAMLC0A0CM, BAMLH0A0HYM2), the SLOOS tightening series (DRTSCILM), the H.8 bank credit, assets and residual series (TOTBKCR, TLAACBW027SBOG, RALACBW027SBOG) and the Z.1 NYSE composite level (BOGZ1FL073164003Q). The Wilshire 5000 (WILL5000INDFC) returned HTTP 404 and is no longer distributed by FRED. Two restrictions matter for the paper: since April 2026 FRED serves only three years of the ICE BofA indices ("Copyrighted: Pre-Approval Required"), so the 2008 and 2020

spread peaks are not in our files and appear in the ledger only as readings of the chart descriptions in the Federal Reserve's November 2020 *Financial Stability Report*; and the S&P 500 level is limited to ten years, which is why long-run equity moments come from the Shiller file.

Distributional Financial Accounts. https://www.federalreserve.gov/releases/z1/dataviz/download/zips/dfa.zip (895,897 bytes, archive dated 2026-09-15); 26 CSV files and three data dictionaries were extracted, covering 1989:Q3–2026:Q2 for the groups TopPt1, RemainingTop1, Next9, Next40 and Bottom50. Only `dfa-networth-shares.csv` is used.

Fernald TFP. https://www.frbsf.org/wp-content/uploads/quarterly_tfp.xlsx (252,789 bytes, vintage "Produced on October 01, 2026"), sheets `annual` (1947–2025) and `quarterly` (1947:Q1–2026:Q2); columns `alpha`, `dtfp` and `dtfp_util`, in 100× (annual) or 400× (quarterly) log changes. The author asks that FRBSF Working Paper 2012-19 be cited.

Shiller data. The current `ie_data.xls` linked from https://shillerdata.com/ (1,675,264 bytes; monthly 1871-01 to 2026-10, with the October 2026 price being the 2 October close and the last two CPI values estimated by the file's own note). Columns used: price, dividend, CPI and the ten-year yield; annual real total returns compound monthly price plus one twelfth of the dividend. The legacy Yale copy ends in 2023-09 and was not used.

SEC filings. Capital expenditure is the XBRL concept `PaymentsToAcquirePropertyPlantAndEquipment` (namespace `us-gaap`) from the SEC XBRL company-concept API (https://data.sec.gov/api/xbrl/companyconcept/CIK<cik>/us-gaap/PaymentsToAcquirePropertyPlantAndEquipment.json, User-Agent header required) for Microsoft, Alphabet, Amazon and Meta, supplemented by Amazon's Q4 2025 earnings release (purchases of property and equipment $131,819 million, twelve months 2025) and Meta's FY2025 exhibit 99.1 ($72.22 billion including finance-lease principal). Microsoft's calendar-2025 figure is derived as the fiscal-2025 total minus the three fiscal quarters reported in 10-Qs, plus the two subsequent quarters ($16.745 + 17.079 + 19.394 + 29.876 = 83.094$ billion); it is the only derived component of the $376.1 billion total.

EIA and open working papers. Electric Power Monthly Table 5.6.A (https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a) is saved as HTML; 2026 values are preliminary. The Flodén–Lindé IIES Seminar Paper 654 (1998) and Gorton–Metrick NBER WP 15223 (2009) PDFs are stored in `data/raw/papers/` and were read for Table 4 and for the haircut-index sentence respectively.

Not obtained. WID.world US wealth shares (client-side charts, no CSV endpoint), SCF 2022 microdata (the bulletin was read for median and mean net worth but publishes no Gini), the Census BTOS data page (JavaScript only; values taken from the Census story of 26 May 2026), the FDIC Quarterly Banking Profile equity ratio (not on the summary page; the H.8 book ratio is used instead), the Cooley–Prescott (1995) chapter text (HTTP 403), the Storesletten–Telmer–Yaron full text (HTTP 404; abstract only) and the OpenAI pricing page (HTTP 403; the developer-docs mirror was read). Each omission is recorded in the manifest and marked NOT OBTAINED in the ledger rather than filled from memory.

E. Replication instructions

The replication package is the git repository `wp25_uqo` (UQO internal git; a public mirror will accompany the journal version). Layout: `src/age/{economy,agents,equilibrium,markets,banking,llm,experi` `scripts/` (numbered pipeline), `tests/{unit,integration,property}`, `docs/` (literature matrix, model specification, parameter ledger, research log, limitations), `experiments/pre_registration.md`, `results/`, `paper/`.

Environment. Python $\geq$ 3.11 with numpy, scipy, pandas, matplotlib, sympy, jsonschema, numba, pyyaml, httpx; pytest and hypothesis for tests. `pip install -e .` installs the package; `requirements.lock.txt` pins the versions used for the paper (Python 3.14.4, numpy 2.4.4, scipy 1.17.1, numba 0.65.1 on macOS/arm64).

Pipeline.

```
python3 -m pytest tests -q # 30+ tests incl. property-based ledger and zero-wedge tests
python3 -I scripts/01_download_data.py # FRED/BEA/Fed DFA/Fernald/Shiller/SEC/EIA
    snapshots -> data/raw
python3 -I scripts/02_build_calibration_targets.py
python3 scripts/00_lq_derivations.py # symbolic derivations (sympy)
python3 scripts/10_model_a.py # Model A results and figures
python3 scripts/20_model_b.py # Model B scenarios (about 5 minutes)
python3 scripts/30_model_c.py # Model C factorial + stress (20 seeds)
scripts/llm_serve.sh start; scripts/llm_tunnel.sh # local inference servers (optional,
    for the LLM lab)
python3 scripts/35_llm_lab.py # LLM laboratory (replays from results/llm_cache if present
    )
python3 scripts/40_figures.py; python3 scripts/50_tables.py
cd paper && make # latexmk -> main.pdf
```

Determinism. All random draws are seeded; Model C uses common random numbers across cells; LLM calls are cached content-addressed so that the LLM results replay exactly from `results/llm_cache` without any model server. The audit logs in `results/llm_audit` record every call. Each experiment writes a manifest with the git revision and package versions.

Tests of the accounting identities. `tests/property/test_ledger.py` applies 100 random sequences of up to 40 operations (dividends, interest, borrowing, repayment, trades, mark-to-market) and asserts after each operation that holdings sum to supply, that the bank's balance sheet balances and that reserves change only through injections; it also checks that defaults and bail-ins are explicit. `tests/unit/test_tasks.py` checks Euler's theorem on factor payments; `tests/unit/test_lq_map.py` checks the goods-market identity of the steady state.

Data. Raw files are not redistributed; `data/raw/README.md` and `data/raw/MANIFEST.json` (URL, timestamp, bytes, sha256) allow re-downloading the same snapshots, subject to the providers' terms.

F. Additional figures and tables

Table F1. Model B: channel decomposition (changes relative to the no-AI economy)

Channel (scenario)	ΔY (%)	Δ labour share	Δ Gini	Δ top 10%	Δr (pp)
AI as production input	97.6	-0.225	-0.098	-0.078	-1.86
Delegation only	3.1	+0.000	+0.018	+0.009	-0.59
Both, broad ownership	107.9	-0.227	-0.108	-0.096	-2.53
Both, concentrated (top 10%)	73.6	-0.217	-0.022	+0.017	-0.05
Scarce compute, broad	36.7	-0.103	-0.049	-0.050	-1.98
Scarce compute, concentrated	24.7	-0.103	-0.026	-0.053	-0.36
Both, broad, no borrowing	108.1	-0.227	-0.121	-0.099	-2.54
Both, broad, cheap fee	109.7	-0.228	-0.120	-0.099	-2.63
Both, broad, no human mistakes	111.0	-0.229	-0.123	-0.102	-2.66

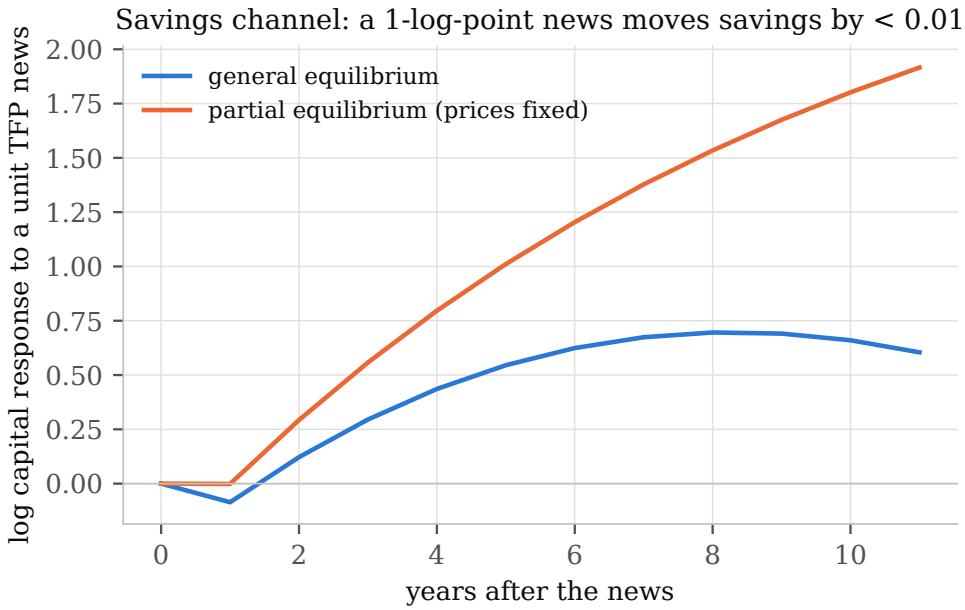


Figure F1. Savings mapping (negative result). Response of log capital to a one-log-point TFP news shock in partial equilibrium (prices fixed) and in the symmetric rational-expectations equilibrium of the log-linearised economy ($\gamma = 2$, $\rho_z = 0.9$, $\delta = 0.06$, $\beta = 0.96$). The partial-equilibrium response, κ_0, is -0.0015: wealth and substitution effects nearly cancel, so the consumption–savings decision carries a negligible stake for delegation. Source: `scripts/10_model_a.py`.

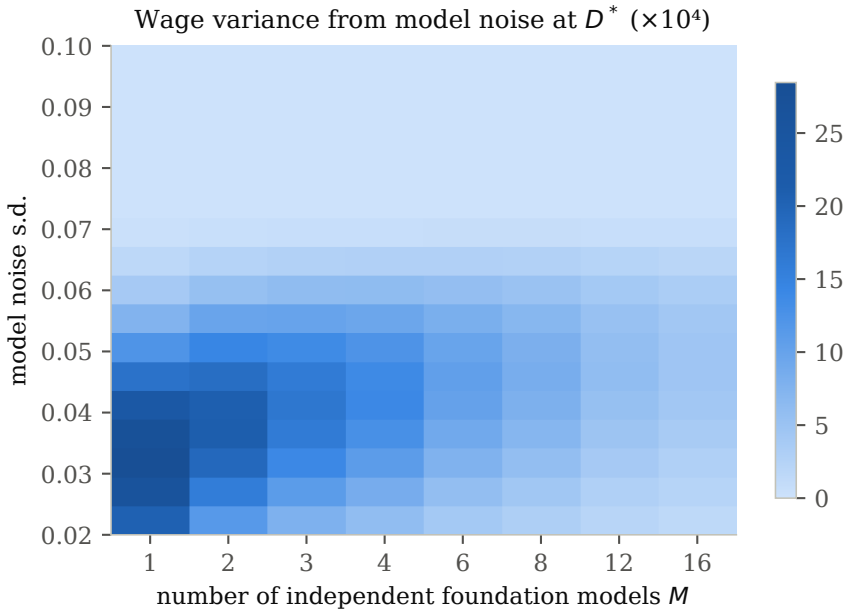


Figure F2. Sensitivity: variance of the log wage attributable to model noise at the equilibrium delegation share, over the model-noise standard deviation (rows) and the effective number of independent foundation models (columns). Source: `scripts/10_model_a.py`.

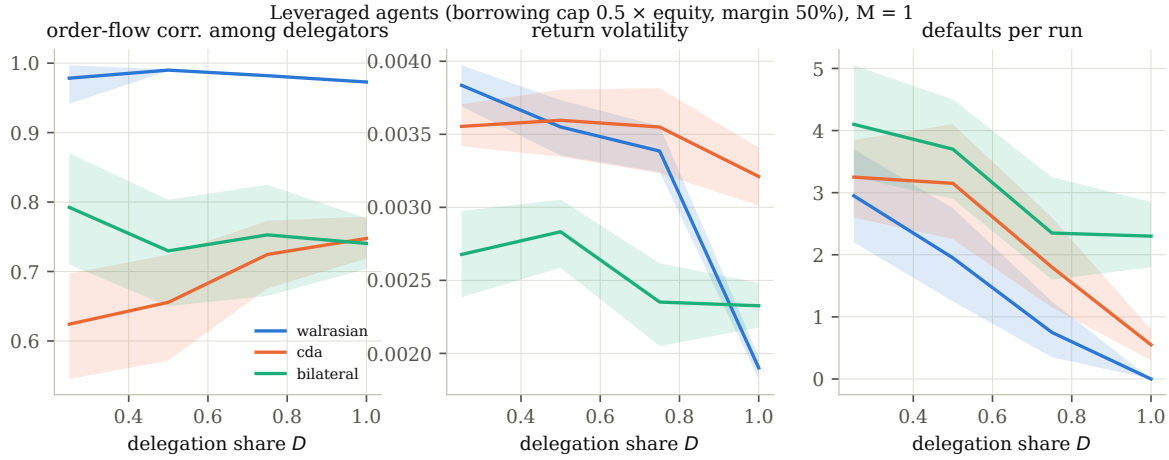


Figure F3. Model C, Bayesian control: order-flow correlation among delegators and return volatility by market design, delegation share and number of models (20 seeds, bootstrap 95% intervals). Source: `scripts/30_model_c.py`.