

SIMS: Language-Model Households and Institutional Regimes in a Demand-Led Macroeconomic Simulation

Paul Kaufmann Josef Menter Jan Schaeffer* Niclas Schult
Franca Wedekind

Bundesdruckerei GmbH, Kommandantenstraße 18, 10969 Berlin, Germany

Preliminary Version: Please do not cite or circulate without permission.

September 29, 2026

Abstract

How do market and accounting constraints shape the behavior of adaptive agents in a language-model-augmented macroeconomic simulation, and what demand-led dynamics emerge from their interaction? Agent-based macroeconomic models in the Post-Keynesian tradition specify household behavior through predefined rules; recent language-model simulations make behavior adaptive but leave the institutional environment fixed. This paper holds a language-model household specification constant while varying the institutional environment. The setting is SIMS, the Simulation of Intelligent Macroeconomic Systems, in which households choose consumption and contract acceptance through a language model, while production, pricing, contracts, credit, fiscal transfers, and market clearing follow explicit rules subject to post-round accounting checks. Four arrangements of contract, production-target, and pricing rules are each evaluated over three runs in an economy of twelve households, three firms, and forty monthly rounds.

Controlled micro screens show that consumption choices differ systematically across personas and decline with lower liquid balances. The macro comparison distinguishes three regimes: persistent slack with incomplete goods clearing and mild deflation when capacity exceeds demand; goods-market clearing at a mechanically low level of activity under a binding production target; and a cumulative price-output collapse when the same target is combined with actual-cost rather than normal-capacity pricing.

Keywords: agent-based modeling; large language models; macroeconomic simulation; Post-Keynesian economics; heterogeneous households

*Corresponding author: jan.schaeffer@bdr.de

1 Introduction

In conventional agent-based models, household demand is specified by a consumption or labor-supply function. Large-language model (LLM) agents offer a different form of specification: a household can receive a character description, current economic conditions, and recent events, then return a constrained choice and a declared rationale. The model, prompt, memory design, and sampling configuration thereby form a household treatment that must itself be characterized before it is used in a macroeconomic comparison.

This raises two questions: First, does the selected language-model household specification behave consistently in controlled settings: do personas differ, and does spending respond to available cash? Second, when that specification is held fixed, how do alternative contract, production, and pricing rules turn the same household demand into different aggregate outcomes? The focus is therefore an application in demand-led macroeconomics: how institutional arrangements shape the adjustment of output, prices, and employment in a small synthetic economy.

We address these questions with the Simulation of Intelligent Macroeconomic Systems (SIMS) framework, a hybrid agent-based model in which households make consumption and labor supply decisions through a LLM, while firms, the bank, government, and markets operate through explicit rules. The hybrid architecture allows controlled separation at the level of experimental design: the model, prompt, population, and sampling configuration remain constant while institutional arrangements are varied.

The study is carried out in two steps. First, controlled micro screens characterize the fixed household specification through persona variation and liquidity responsiveness. Second, a macro evaluation compares four institutional arrangements over three independent runs each, in a twelve-household, three-firm economy over forty monthly rounds. This comparison yields three distinct macroeconomic regimes: persistent slack with incomplete goods clearing, goods-market clearing at a mechanically low level of activity, and a cumulative price-output collapse under actual-cost pricing. These regimes are conditional on the selected household specifications and arrangements; only the cost-basis contrast between the two target-limited arrangements isolates a single rule.

The contribution lies in a controlled hybrid design in which a contextual heterogeneous household treatment remains fixed while explicit institutional arrangements vary. This permits arrangement-level differences in the reported macro paths to be interpreted against explicit demand-led mecha-

nisms, while preserving a decision record for later household-level analysis. The present comparison does not use declared rationales to explain its aggregate results. [Section 2](#) positions this design in the relevant literature. [Section 3](#) describes the economy and its recorded decisions, and [Section 4](#) provides identification. [Section 5](#) and [Section 6](#) present micro and macro evidence. [Section 7](#) provides a comparative interpretation and clarifies what LLM-based households add in the present design. [Section 8](#) concludes the paper.

2 Related Work and Positioning

Rule-based macroeconomic agent-based models ([Tefatsion and Judd, 2006](#)) make the behavioral and institutional rules that organize an economy explicit. They have been used to study the interaction of demand, production, finance, unemployment, and macroeconomic instability ([Dosi et al., 2010](#); [Gualdi et al., 2015](#); [Seppecher, 2012](#)). The post-Keynesian branch of this literature treats effective demand, markup pricing, and endogenous money as the organizing mechanisms and builds stock-flow-consistent benchmark models around them ([Caiani et al., 2016](#); [Hein, 2023](#); [Lavoie, 2014](#)). In these models, a result is interpreted against specified rules for household demand, firm behavior, and market clearing. The present paper remains within that tradition for firms and institutions.

Research using LLMs has taken two related but distinct paths. [Filippas et al. \(2024\)](#) use language models as simulated economic subjects in controlled choice settings. [Aher et al. \(2023\)](#) introduce Turing experiments that test whether language models reproduce findings from human-subject studies, including an ultimatum game; they do not construct an economic simulation. The EconAgent framework ([Li et al., 2024](#)), by contrast, places language-model agents in a macroeconomic environment in which they make work and consumption decisions using character and memory modules. More recently, [Liu et al. \(2026\)](#) explicitly assess language-model households in a macroeconomic agent-based model at both the micro and the macro level. [Dwarakanath et al. \(2024\)](#) provide a multi-agent simulator in which households and firms follow fixed rules or learned policies alongside a rule-based central bank and government; their household treatment is a learned policy rather than a characterized persona specification.

What these studies share is that the institutional environment is a fixed backdrop: market, pricing, and contract rules are not varied as a treatment. Where the household treatment is

characterized at the micro level, as in [Liu et al. \(2026\)](#), it is then not held fixed across alternative institutional rules. Taken together, this literature motivates separate examination of the responses generated by a particular household specification. Nominally different persona descriptions may still produce a common model-default response, and a declared rationale need not predict the recorded choice. The micro screens in [Section 5](#) therefore test whether the selected specification produces differentiated responses to persona and liquidity conditions.

In SIMS, firms, markets, credit, and fiscal execution follow explicit rules that define the institutional setting, while households choose consumption and contract acceptance in response to persona and economic situation. Each household decision is recorded with its context, declared rationale, structured choice, and realized transaction. The rationale is not treated as a verified causal account, but it makes the household decision open to examination.

The contribution is therefore a controlled hybrid application. It brings a common, inspectable household treatment into a demand-led macroeconomic comparison under explicit market and accounting institutions. With the household specification held constant, the four configurations permit arrangement-level comparisons of the resulting aggregate paths. The present comparison does not use the declared rationales to explain those paths, and because the linked rules change together through most configurations, it identifies no individual-rule effect beyond the one conditional contrast between the two target-limited arrangements (see [Section 6.4](#)).

3 Model

3.1 Actors, flows, and constraints

SIMS represents a monetary economy in discrete monthly rounds. The simulation distinguishes four types of entities: households, firms, a bank, and a government. Households are workers and consumers. They receive wages when contracted, pay income tax, receive an unemployment benefit when eligible, and allocate part of their cash balance to consumption. Firms apply for credit, invest, hire labor, produce a single consumer good, sell output, and repay their loans. The bank grants loans and receives repayments, while the government collects taxes and pays transfers.

The simulation runs reported in this paper contain twelve households, three firms, one bank, and one government. The twelve households are constructed from a fixed set of synthetic profiles

described in [Section 4.1](#). Age enters the persona description but does not affect eligibility for the common 40-hour labor contract. Each household also carries an energy stock, described in [Section 3.3](#), that is depleted by work and restored by consumption and non-working time.

Household consumption and labor-contract acceptance are language-model decisions. All remaining agents follow explicit rules. This division fixes the production, pricing, credit, fiscal, and market-clearing mechanisms while household choices remain conditional on persona and current economic circumstances. Keeping firms, the bank, and government rule-based also prevents changes in the institutional arrangement from simultaneously changing the household behavioral specification under comparison.

3.2 Within-round sequence

Each household carries a memory document that is rendered into every decision prompt. The memory consists of a fixed identity section generated from the persona, the seven most recent events maintained by the simulation code, and up to seven retained insights. At the start of round t , the event list is updated with events from round $t - 1$; the first round has no preceding events. From the second round onward, a second language model, Qwen3 14B, rewrites the insight list using its previous version and the events of the preceding round. This model runs at temperature zero. If the returned response is malformed, the previous insight list is retained. Household decisions are therefore potentially path-dependent because retained insights can enter later decision prompts.

Firms then apply for credit, the bank evaluates the applications, and approved funds finance investment. Expiring contracts, as well as contracts whose required hours can no longer be covered by the household’s energy stock, are removed. In the first round, before any sales history exists, each firm plans for 40% of full-capacity output and hires two workers. Employment therefore starts at 50% of the twelve-household population and subsequently adjusts as contracts expire.

Uncontracted households next decide whether to accept the offered 40-hour weekly contract. This labor-supply decision occurs before firms set their production target and resulting demand for additional workers. The common nominal wage adjusts upward under excess labor demand and downward under excess supply, after which labor-market clearing matches accepting households to vacancies and wages are paid.

The fiscal stage collects household and corporate taxes and then pays transfers. Households

subsequently choose their desired nominal expenditure. The goods market converts the capital stock and contracted labor into output, matches household budgets with available goods, executes purchases, records each firm’s unsold share, and updates prices and firms’ demand signals. Goods are perishable: output not sold within the round is discarded. The unsold share is therefore a flow measure of excess supply rather than an inventory stock. Bank settlement then collects repayments, pays interest on positive balances, and determines whether each firm remains operational in the next round. Finally, post-round accounting checks the resulting transactions and balance sheets.

3.3 Language-model households

Each round, uncontracted households decide whether to accept the standard labor contract. Before goods-market clearing, every household also chooses desired nominal expenditure $d_{h,t}$, subject to its available cash balance $b_{h,t}$:

$$0 \leq d_{h,t} \leq b_{h,t}. \tag{1}$$

Realized expenditure can be lower than $d_{h,t}$ when insufficient goods are available. Gemma 4 31B generates both contract-acceptance and expenditure choices from the household’s persona, cash balance, market conditions, and memory. Decisions are sampled at temperature 0.7, so repeated decisions in the same economic situation may differ. The controlled micro screens in [Section 5](#) characterize this variation and test whether expenditure responds systematically to persona and liquidity.

Households additionally carry an energy stock bounded between 0 and 100 that summarizes their capacity to work. Each month, energy falls by a fixed base cost and by a constant amount per hour worked. It recovers by a constant amount per hour not worked and by a concave function of units consumed, so the first units of consumption restore more energy than later units. A household may accept the standard contract only if its current energy is sufficient to cover the contractual hours for the month, and an existing contract is terminated when this condition ceases to hold. Energy therefore links past work and consumption to subsequent labor supply. In the reported runs, however, mean energy remains between 83 and 98, and no household is excluded from work because of insufficient energy. The relevant parameter values are reported in [Appendix A](#).

3.4 Firms, prices, credit, and fiscal policy

Potential output of firm i is determined by a two-factor Cobb–Douglas production function,

$$Q_{i,t}^{pot} = AK_{i,t}^\beta L_{i,t}^\alpha, \quad (2)$$

where A denotes total factor productivity, $K_{i,t}$ the firm’s capital stock, $L_{i,t}$ its monthly labor hours, and α and β the labor and capital elasticities. The parameterization satisfies $\alpha + \beta = 0.9$, implying mildly decreasing returns to scale. Firms form an output target $\bar{Q}_{i,t}$ from recent demand, sales, and unsold-share signals. In configurations with a binding output constraint, realized production is

$$Q_{i,t} = \min \left\{ Q_{i,t}^{pot}, \bar{Q}_{i,t} \right\}. \quad (3)$$

Otherwise, realized production equals $Q_{i,t}^{pot}$. Under target limitation, a firm can therefore produce less than its capital stock and contracted labor would technically permit, leaving part of both productive inputs idle.

The cost basis used for price bounds is wage cost $W_{i,t}$ divided either by realized output or, when the normal-capacity rule applies under a binding target, by potential output:

$$c_{i,t}^A = \frac{W_{i,t}}{Q_{i,t}}, \quad c_{i,t}^N = \frac{W_{i,t}}{Q_{i,t}^{pot}}. \quad (4)$$

Firm prices remain within fixed lower and upper markup bounds around the applicable cost basis. A low unsold share generates a signal for a possible price increase, while a high unsold share generates a signal for a possible reduction. Independently of this signal, a firm leaves its price unchanged with probability one half in a given month, introducing Calvo-type nominal stickiness (Calvo, 1983). The distinction between actual and normal-capacity cost is central to the institutional comparison: when an output target binds, the actual-cost rule raises unit cost as $Q_{i,t}$ falls, whereas the normal-capacity rule does not.

Credit applications are rejected if the borrower has previously defaulted, if the requested loan

exceeds the bank’s reserve-constrained lending capacity, or, when positive monthly revenue has been recorded, if it exceeds a revenue-based ceiling. For requested borrowing $\ell_{i,t}$, bank cash B_t , reserve ratio ρ , firm revenue $R_{i,t}$, and revenue multiple μ , the two quantitative constraints are

$$\ell_{i,t} \leq (1 - \rho)B_t, \quad \ell_{i,t} \leq \mu R_{i,t}. \quad (5)$$

No loan default occurs in the reported runs.

The fiscal rule levies flat taxes on household income and positive firm profits. For household income $Y_{h,t}$ and firm profit $\Pi_{i,t}$, assessed taxes are $\tau_h Y_{h,t}$ and $\tau_f \max\{\Pi_{i,t}, 0\}$, respectively, subject to the payer’s available cash. Unemployed households receive a benefit indexed to the common wage rate between a lower and an upper bound. These bounds are set equal in all reported runs, making the benefit a fixed \$150 per month rather than allowing it to decline with wages during deflation. The government cannot run a deficit: if available cash above its reserve is insufficient, transfers are scaled down proportionally.

Fixed parameter values and configuration-specific parameter changes are reported in [Appendix A](#). The complete equations for the mechanisms summarized here are provided in [Appendix B](#).

3.5 Accounting and recorded decisions

All monetary flows are recorded as transactions. After every round, the framework verifies that each goods purchase has a corresponding firm revenue entry, that firm debt equals the bank’s outstanding loan stock, that ordinary agents do not hold negative cash balances, and that bank cash remains non-negative. These checks impose accounting consistency on the realized transactions independently of the household decisions generated by the language model.

Across the twelve 40-month configuration runs, the available records contain each household’s stated rationale, intended expenditure, and realized purchase outcome for every month. When a purchase occurs, it can be linked to the corresponding transaction. The archived records do not, however, retain the complete rendered prompt or contemporaneous memory document. They therefore permit a descriptive comparison of individual decisions and realized outcomes, but they do not support the claim that stated rationales causally generated the aggregate mechanisms analyzed

below. The declared rationales are consequently not used to explain the macroeconomic results in this paper; the complete decision records are retained as the basis for subsequent household-level analysis.

4 Experimental Design

4.1 Experimental strategy and fixed household treatment

The study proceeds in two steps. First, controlled micro screens characterize the selected household treatment through persona variation and liquidity responsiveness. Second, the macro evaluation holds this treatment fixed while varying the institutional environment. The macro comparison therefore does not compare alternative household models; it examines how a common adaptive household treatment interacts with different combinations of firm, contract, and pricing rules.

The completed four arrangements use the same twelve synthetic personas, decision prompt, primary decision model, memory mechanism, and sampling settings. Contract acceptance and consumption use Gemma 4 31B at temperature 0.7. Qwen3 14B at temperature zero consolidated household memory. The household treatment is thus not varied across arrangements. It is not a treatment-level explanation for systematic differences among the institutional arrangements.

The personas are drawn from the synthetic subset of the MatrAIx Persona 1M dataset (Li et al., 2026; MatrAIx, 2026). Each persona describes relatively stable demographic and preference attributes, including economic orientation, risk attitude, financial priorities, and decision style. These descriptions are distinct from the household’s simulated economic state: current cash balances, employment status, wages, and other contemporaneous conditions are supplied by the model in each round. The twelve personas were selected to span a range of economic orientations and risk attitudes rather than to represent an empirical population distribution. Section 5 tests whether these persona descriptions generate systematic differences in household decisions under controlled conditions; Appendix E documents the retained fields and provides an example profile.

4.2 Institutional arrangements and identification

The macro experiment compares four institutional arrangements. Each arrangement is a fixed combination of contract duration, production-target rules, pricing rules, and their associated ad-

justment parameters. Throughout the paper, an *arrangement* refers to this institutional treatment, while a *regime* refers to the aggregate macroeconomic path that emerges from it.

Most arrangements modify several institutional elements jointly. Contract duration, target limitation, and the cost basis used in price adjustment change together across much of the comparison; the volatile-activity arrangement additionally changes the markup bounds, wage-adjustment speed, and production-signal parameters. Differences between these arrangements therefore identify the consequences of complete institutional configurations rather than the causal effect of an individual rule.

One comparison is more tightly identified. The two target-limited arrangements share the same monthly contracts and production-target rule and differ only in whether price bounds are based on actual-average cost or normal-capacity cost. Their contrast therefore isolates the cost-basis rule conditional on the target-limited, monthly-contract setting. No corresponding single-rule interpretation is available for the remaining arrangement comparisons.

This design conditions the macro results on the selected household treatment. Because its components are not varied across arrangements, systematic contrasts cannot be attributed to treatment differences. Conversely, the experiment does not establish that a different household treatment would preserve the observed ordering or qualitative differences across institutional arrangements.

4.3 Outcomes, scale, and repetition

The main macroeconomic outcomes are employment, the goods-clearing rate, price growth and price volatility, and smoothed output growth. Employment and price measures are reported over rounds 6–40, leaving the first five rounds as a common warm-up period. Late employment and goods clearing are measured over rounds 36–40. Smoothed output growth is calculated over rounds 20–40. The exact aggregation conventions for these measures are defined in [Appendix C](#). No arrangement is treated as an empirically validated baseline.

The completed comparison contains twelve households and three firms across four institutional arrangements, each evaluated over three stochastic runs of forty monthly rounds. A common seed aligns the rule-based sources of randomness across arrangements, including goods-trading order, the matching of accepting households to firms, price-adjustment magnitudes, and the Calvo draws governing whether a signal-driven price adjustment is applied. The same seed is passed to the

language-model sampler. Language-model decisions are not exactly reproducible across executions, however, so an identical seed and configuration can still yield different household choices. The seeds therefore align rule-based stochasticity and distinguish repeated simulation paths rather than making individual runs bit-reproducible. The simulation run, rather than a monthly observation, is the unit of evidence.

Three runs document material run-level variation but do not support precise dispersion estimates, formal significance tests, or population inference. Arrangement-level medians summarize the three runs, while seed-level outcomes are reported where variation across runs is substantively relevant. At this population size, one active contract corresponds to 8.3 percentage points of employment, so persistent differences smaller than one contract are not interpreted as distinct labor-market adjustment.

The small economy supports a qualitative comparison of economic mechanisms, not population inference or precise statistical effect estimates. [Section F.1](#) reports a supplementary robustness check with an independently drawn population of 24 households and six firms. It is not used as a basis for the main claims and, because it changes population scale and persona composition together, does not isolate the effect of scale alone.

5 Results: Micro-level Evidence

5.1 What the screens establish

The micro screens ask two questions: whether the supplied persona descriptions produce different spending choices in a common economic situation, and whether lower liquid balances reduce spending. However, they do not validate the households against external data or demonstrate that these responses are necessary for the macro regimes.

Each screen presents a fixed economic situation to a household and records its chosen expenditure. The sample consists of the twelve synthetic MatrAIX personas described in [Section 4.1](#) (compare with [Appendix E](#)). An otherwise identical no-persona condition provides a reference point. Amounts are expressed in the framework’s unit of account, denoted by \$. The common household specification is described in [Section 3](#) and [Section 4](#).

5.2 Persona differentiation

At the common cash balance of \$500, each persona made twenty spending decisions. Average spending across the twelve personas was \$78.6, compared with \$41.6 in the otherwise identical no-persona condition. Persona mean expenditures ranged from \$25 for Daniel to \$170 for Ruth. Daniel’s profile emphasizes wealth preservation and skepticism about consumption, whereas Ruth’s is risk-seeking and pro-consumption. The direction of the two means is consistent with their descriptions, without identifying any one attribute as its cause. [Figure 1](#) displays the full spread and its within-persona variation. [Appendix D](#) reproduces one rendered prompt and response to document the decision-level material underlying the screen; it is illustrative, not additional evidence.

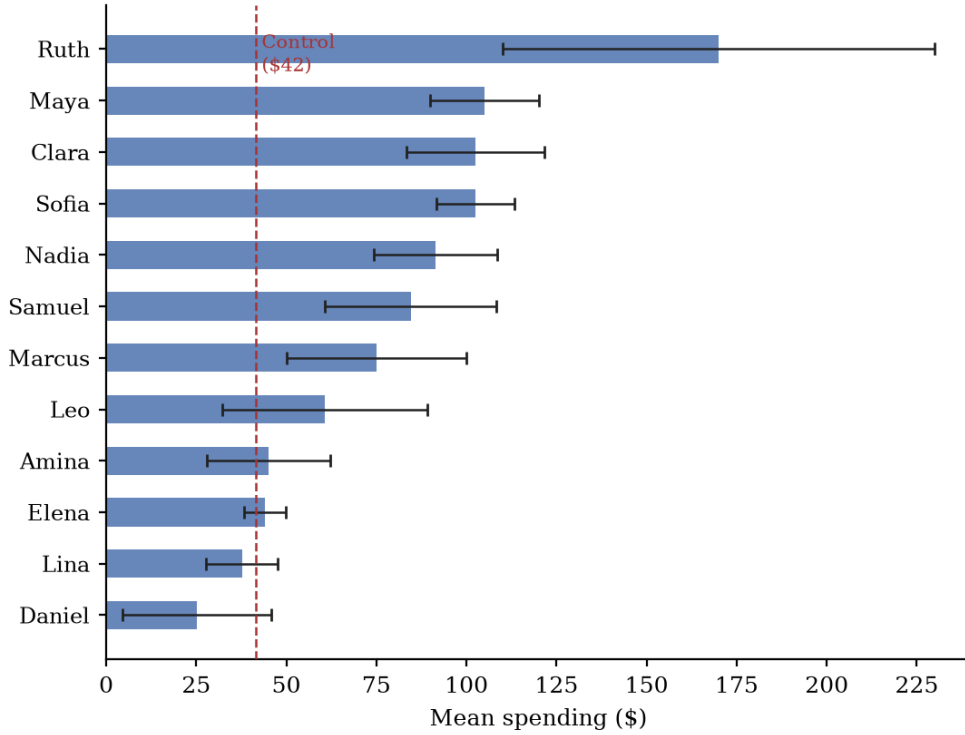


Figure 1: Mean spending by persona at a common balance of \$500. Error bars show one standard deviation across twenty decisions. The dashed line is the otherwise identical no-persona reference.

The intraclass correlation ($ICC(1,1)$) is 0.705. In other words, approximately 70% of the observed variance is attributable to between-persona variation in this screen ([McGraw and Wong, 1996](#)). The underlying F statistic is 48.8, with 11 and 228 degrees of freedom ($p < 0.001$). A second execution of the same screen with identical settings and seeds yields an ICC of 0.700 and a mean of \$79.9. Language-model sampling is not fully reproducible across executions (see [Section 4.3](#)), so this replication serves as a test of sampling reproducibility rather than a duplicate.

This demonstrates that the differentiation of personas within the selected set survives repeated sampling. It does not ascribe the difference to one profile attribute or generalize it to a representative population.

5.3 Response to liquid balances

The balance screen holds the persona and other stated conditions fixed while varying the cash balance. The baseline uses twenty decisions per persona; each lower-balance condition uses ten. [Table 1](#) reports the resulting means. Each test pairs one mean per persona under the baseline and alternative balance, so it has eleven degrees of freedom. The reported t -statistics test the positive difference between baseline spending and spending at a lower balance.

Table 1: Liquidity screen

Cash balance	Mean spending	Spending reduction relative to baseline	t	p
\$500	\$78.6	—	—	—
\$250	\$45.6	\$33.0	4.82	< 0.001
\$100	\$24.0	\$54.7 ¹	6.83	< 0.001

¹ Differences use unrounded persona means; displayed spending means are rounded to one decimal place.

Spending declines for eleven of twelve personas at \$250 and for all twelve at \$100; in a repeated execution of the \$250 screen, all twelve decline, with a mean reduction of \$31.4 and $t(11) = 4.84$, $p < 0.001$. The implied marginal propensity to consume from liquid balances is 0.132 over the \$500-to-\$250 change and 0.137 over the \$500-to-\$100 change: about 0.13 across the two balance changes. This number summarizes the response in this static screen; it is not an estimated structural consumption parameter. The shared persona template and economic context used in these screens are reproduced in [Appendix D](#).

The direction of the response is consistent with the basic intuition of buffer-stock saving: households with less readily available cash restrain current purchases to preserve a financial buffer ([Carroll, 1997, 2001](#)). The screen does not, however, observe a household building a buffer over time and is therefore not a direct test of that dynamic theory.

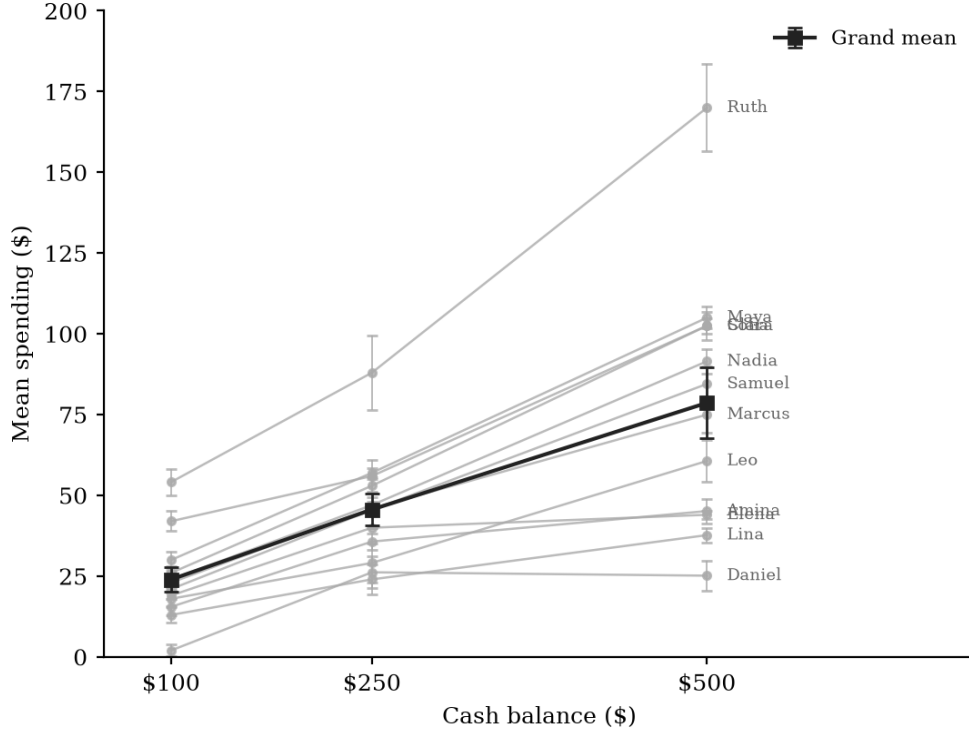


Figure 2: Mean spending by persona at balances of \$100, \$250, and \$500. Gray lines show individual personas; the black line is their mean. Higher-spending personas make the largest absolute reductions when liquidity tightens.

5.4 Summary

The screens establish two limited properties of the fixed household specification: persona descriptions differentiate spending under a common state, and lower liquid balances reduce spending in the supplied context. These are the micro-level properties used to interpret the demand side of the subsequent macro comparison.

Two further screens, run with the same design and ten decisions per persona, are reported here only as limits of the test. Halving or doubling the goods price left mean nominal spending unchanged (\$80.1 and \$75.9 against \$78.6; $|t|$ at most 1.05), and lowering household energy to 50 or 25 of 100 left it unchanged as well ($|t|$ at most 1.64). This can likely be attributed to the prompt stating the price without relating it to purchasing power, and stating the energy level and that consumption restores it without attaching any consequence of low energy for work or income. Their one consequence for [Section 6](#) is that, in this static screen and within this specification, a higher price reduces the quantity a household buys roughly one for one. The screens therefore do not establish a general price elasticity, an energy-response mechanism, or external behavioral validity.

6 Results: Macro-level Evidence

6.1 The macro comparison

[Section 5](#) characterizes the household specification used in the simulations. This section asks a different question: how does the same household treatment play out when firms and markets operate under four alternative institutional arrangements? The population, household decision model, and decision prompt are held constant. Differences in the reported paths therefore describe the joint operation of the rules in each arrangement, not differences in the household specification.

Each arrangement is run with the same three seeds, resulting in twelve completed simulation paths in a twelve-household, three-firm economy over forty monthly rounds. The repetitions show whether a regime recurs beyond one stochastic path; three runs do not support formal inference or estimates of a population distribution. The arrangements were fixed before the runs. They are not a calibration exercise and, with the one conditional exception discussed in [Section 6.4](#), they do not identify the effect of any individual rule.

The argument proceeds in three steps. [Section 6.2](#) clarifies the differences between the four arrangements and explains why they form a useful comparison. [Section 6.3](#) then identifies the outcomes that differ across them. Finally, [Section 6.4](#) reads each regime through the economic sequence of available output, goods clearing, prices, and employment.

6.2 Four macro configurations

The comparison is not a factorial sweep of all possible firm rules. It begins with an unconstrained production case in which unsold output recurs month after month. Two target-limited arrangements then hold the monthly-contract setting in common while using alternative cost bases for pricing. This makes the consequence of pricing against potential rather than realized output visible within that broader institutional setting. The fourth, volatile-activity arrangement is a secondary contrast: it retains unconstrained production and longer contracts, but changes the markup bounds, the wage-adjustment speed, and the production-signal parameters together.

Target limitation determines whether production can be held below the output implied by existing labor and capital. Contract duration determines how often firms can revise the number of workers they retain. The price basis matters when output is target-limited, because payroll costs

per realized unit then differ from costs calculated at normal capacity. The configurations combine these rules, as shown in [Table 2](#); they should consequently be read as alternative arrangements, not as isolated policy interventions. [Section A.2](#) reports their configuration-specific parameter values.

Table 2: Institutional arrangements in the macro comparison (parameter values in [Section A.2](#))

Configuration	Target limit	Contract duration	Cost basis of the floor	Further differences
Persistent slack	No	4 months	Actual cost (equals normal capacity)	—
Underemployed market clearing	Yes	1 month	Normal-capacity cost	—
Unstable monthly adjustment	Yes	1 month	Actual-average cost	—
Volatile activity	No	4 months	Actual cost (equals normal capacity)	Markup band, wage and production adjustment, production-signal rules

6.3 Results

All twelve planned arrangement–seed combinations completed. Two of them (persistent slack, seed 42; unstable monthly adjustment, seed 43) completed on a second attempt after an interrupted first run; the interrupted attempts are excluded. Run metadata are comprehensively documented in the evidence records. [Table 3](#) reports the median of the three runs for each arrangement. The outcome definitions and aggregation conventions are given in [Appendix C](#).

The principal differences concern goods clearing and prices. The underemployed-market-clearing arrangement has a late clearing rate near 88%; persistent slack and volatile activity remain near 35–40%. The unstable-monthly-adjustment arrangement instead stands out due to price growth above 30% per round and much greater price volatility. These differences are interpreted below.

Employment does not discriminate among the four arrangements at this scale. Across the twelve runs, mean employment ranges from 23.8% to 27.1%. The widest difference between two runs therefore remains below half of one active contract on average over the evaluation window.

Table 3: Macro outcomes: medians of three seed-specific runs

Configuration	Late goods-clearing rate (rounds 36–40)	Mean price growth (% per round, rounds 6–40)	Inflation volatility (percentage points, rounds 6–40)	Smoothed output growth (% per round, rounds 20–40)	Mean employment (rounds 6–40)	Late employment (rounds 36–40)
Persistent slack	35.1%	−1.21%	0.87	+0.20%	25.0%	25.0%
Underemployed market clearing	87.6%	−0.81%	0.28	−0.48%	25.0%	25.0%
Unstable monthly adjustment	57.1% ¹	+32.10%	56.81	−18.28%	24.5%	25.0%
Volatile activity	40.0%	−1.55%	2.54	+0.65%	25.7%	25.0%

¹ Computed on production below one unit per month in rounds 36–40; not economically interpretable (Section 6.4).

The configuration medians are all close to three employed households, and late employment is 25.0% in each median. Figure 3 shows the employment and goods-clearing trajectories.

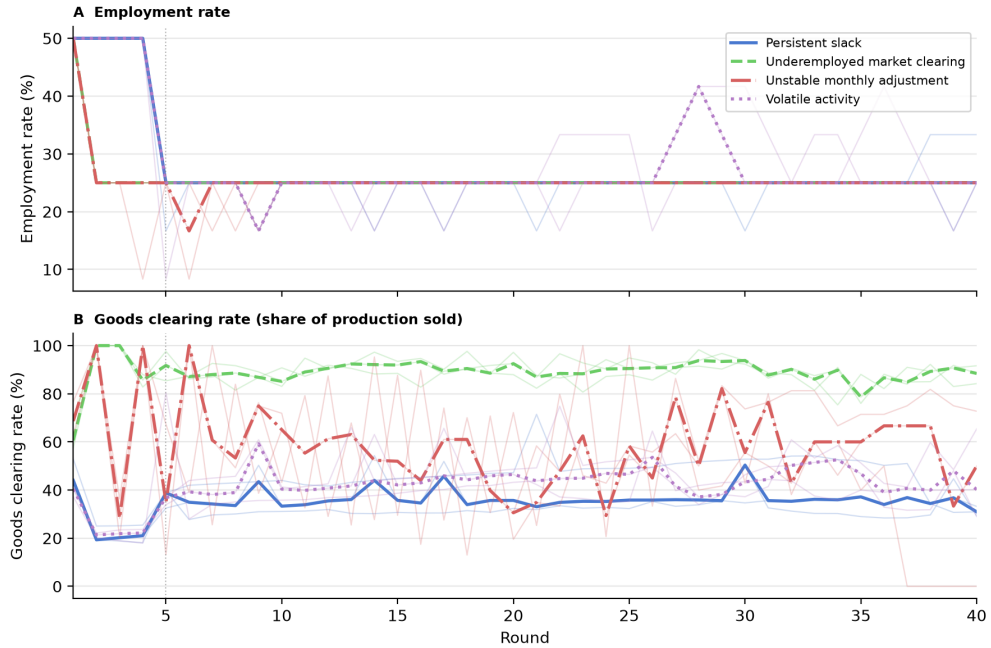


Figure 3: Employment and goods-clearing trajectories by institutional arrangement. Thin lines show the three individual seed paths. Thick lines show the round-by-round median across the three seeds, with a distinct line style per arrangement; these are not themselves simulated paths. Employment starts at 50% because every firm hires two workers in the first round; the two one-month-contract arrangements fall to 25% from round 2 and coincide thereafter, the two four-month-contract arrangements from round 5. The vertical dotted line marks the end of the five-round warm-up.

6.4 Interpreting the four regimes

Persistent slack: unconstrained production does not absorb the demand shortfall. Firms produce from contracted labor and capital without an output ceiling, yet late goods clearing remains between 31.2% and 38.8%. Mild deflation does not eliminate the excess supply; the unsold output is discarded at the end of each month. Production is essentially flat: firms operate their contracted labor and capital at full output in every round (Figure 5), and the small median growth of +0.20% reflects a fourth contract that is filled in some late rounds of one seed. Why does deflation not raise sales? Because prices and nominal incomes fall together: between rounds 6–10 and 36–40, the price level falls by 25–32% and the nominal wage by 27%, so nominal household spending falls by 20–40% while the number of units bought stays within a few percent of its early level (about 135–195 units per month against production of 454). Falling prices leave real purchases approximately unchanged when the income from which they are financed falls at a similar pace. This is stagnation with persistent excess supply rather than recovery or collapse.

Underemployed market clearing: high clearing does not imply high activity. Target limitation reduces the amount offered for sale; a clearing rate of 86–89% is therefore a high ratio of sales to this smaller available output, not evidence of high aggregate demand or income. Firms start with two workers each (Section 3.2); once those one-month contracts expire, the production target is low enough that one worker per firm suffices, so employment settles at three households from the second round onward and output declines mildly. With six firms and 24 households, the same rule employs exactly six (Section F.1). The arrangement reduces unconsumed output; it does not produce a high-employment equilibrium.

Unstable monthly adjustment: a cost floor that follows realized output. This arrangement differs from underemployed market clearing in one rule only: the price floor is computed from actual rather than normal-capacity unit cost (Table 2). Under actual-average-cost pricing, a fall in target-limited output raises payroll cost per realized unit and, with it, the price floor. Higher prices reduce the quantity that a given household budget can purchase; firms lower their targets further, and unit cost rises again. In all three runs, this cumulative process drives economy-wide production below one unit per month between rounds 19 and 22 and keeps it there (Figure 4 and Figure 5). By round 40, the price level stands between roughly 570 and 5,500 times its first-round value. The markup band does not cap this rise, because it is defined relative to unit cost, which grows without bound as production approaches zero while one worker per firm remains paid.

Mean price growth over rounds 6–40 is 27.7–37.1% per round, with a volatility of 54–121 percentage points, far above the other arrangements. In later rounds, three households remain under contract and are paid, out of the firms’ initial cash balances, for output that has effectively ceased; no firm exhausts its cash within forty rounds. Because production has collapsed, the late goods-clearing rate carries no economic information for this arrangement; [Table 3](#) reports it only for completeness. Within the target-limited, monthly-contract setting, the contrast with underemployed market clearing therefore isolates the price–cost basis: normal-capacity costing yields near-clearing at low activity, whereas actual-cost costing yields a cumulative price–output collapse. The comparison does not show how either basis would behave without target limitation or under longer contracts.

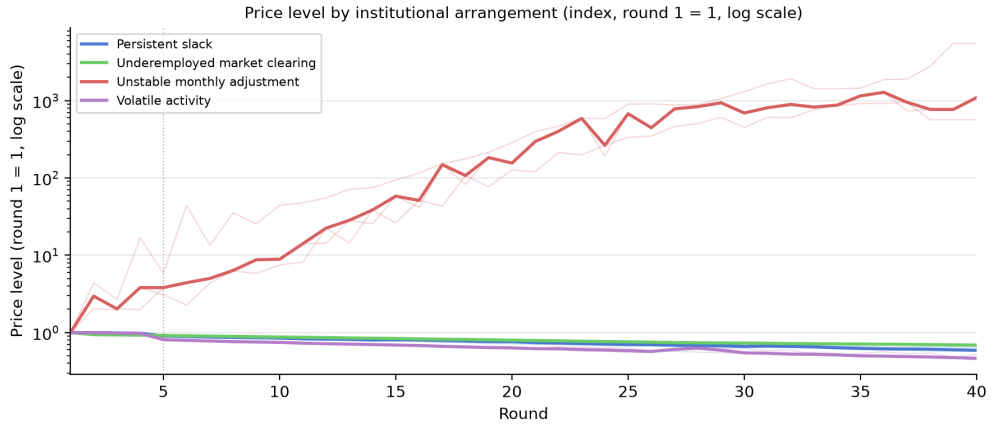


Figure 4: Price-level trajectories by institutional arrangement, with each run indexed to its own first-round price and drawn on a logarithmic scale. Thin lines show the three individual seed paths. Thick lines show the round-by-round median across the three seeds; these are not themselves simulated paths. The vertical dotted line marks the end of the five-round warm-up.

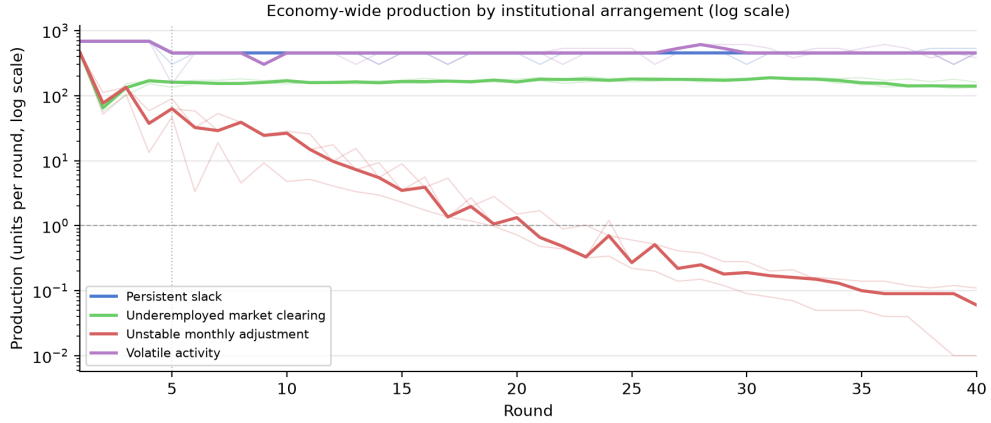


Figure 5: Economy-wide production per round by institutional arrangement on a logarithmic scale. Thin lines show the three individual seed paths; thick lines show the respective round-by-round median across the three seeds. The dashed horizontal line marks one unit per round; the vertical dotted line marks the end of the five-round warm-up. Persistent slack and volatile activity produce at the capacity of their contracted labor, underemployed market clearing at the target-limited level, and unstable monthly adjustment falls below one unit between rounds 19 and 22.

Volatile activity: a volatile variant of persistent slack. This secondary contrast retains unconstrained production and four-month contracts, but changes the markup bounds, wage-adjustment speed, and production-signal rules jointly. Because the target is not binding, firms again produce at the capacity of their contracted labor, and late clearing remains between 34% and 50%, leaving persistent excess supply; smoothed output growth is modestly positive at the median (+0.65%). Price growth is more volatile (2.5–3.9 percentage points against 0.9 under persistent slack) around a similar mild deflation (−1.2% to −1.6% per round); the arrangement’s narrower markup band and faster adjustment rules produce this jointly, and it cannot be concluded from the comparison which of them is responsible. Nothing in the arrangement raises unit cost when sales fall, so no collapse occurs. The employment result for this arrangement is not interpreted separately, because the difference from persistent slack is below one household in this population.

Taken together, the four arrangements produce three distinct goods-market and price regimes under a common household treatment: persistent excess supply, high clearing through reduced output, and a cumulative cost–price collapse. Volatile activity is a variant of the first, differing in price volatility but not in clearing or activity. [Section 7](#) relates these conditional regimes to the paper’s broader theoretical and methodological contribution.

7 Discussion

The experiments hold a selected household specification constant while varying four institutional arrangements defined by firm, contract, and pricing rules. The resulting macroeconomic differences therefore describe how these institutional configurations shape aggregate dynamics under a common household-demand specification. In most comparisons, however, several rules change jointly: target limitation, contract duration, and the price–cost basis are altered together, so the resulting differences cannot be attributed to any one rule in isolation. The exception is the comparison between the two target-limited arrangements, which differ only in their price–cost basis and therefore permit a conditional single-rule contrast within that institutional setting ([Section 6.4](#)).

The micro screens characterize the household specification that is held fixed in these macro comparisons. Under the stated prompt, spending differs systematically across personas and declines as liquidity becomes more constrained. An ICC of 0.705 indicates that most of the observed variation in the screens lies between persona means rather than within personas across repetitions. Because aggregate household demand in the macro simulations is constructed from these individual decisions, the chosen persona composition necessarily contributes to aggregate demand. Since that composition is held fixed throughout the macro experiments, its separate effect is not identified. The micro results should therefore be interpreted as a characterization of the selected household specification, rather than as an external validation of household behavior or as an independent explanation of the observed macroeconomic regimes.

7.1 Interpreting the macro regimes in post-Keynesian terms

The four arrangements can be read against familiar post-Keynesian mechanisms ([Godley and Lavoie, 2007](#); [Kalecki, 1954](#); [Lavoie, 2014](#); [Rowthorn, 1977](#)). This is an interpretive comparison, not an empirical estimate or a causal identification of the theoretical mechanisms.

Persistent slack resembles an underconsumption or stagnation regime. Production is not constrained by a binding output target, yet household demand remains insufficient to absorb the available supply. The mild deflation observed during the simulation does not restore sales. The regime therefore illustrates that weak effective demand need not generate an automatic return to full capacity utilization, and that falling prices may fail to stimulate demand when wages decline alongside them ([Kalecki, 1954](#)).

In the underemployed market-clearing regime, supply is mechanically constrained to the level of weak demand. Goods clear not through a recovery of demand, but because production and employment are restricted by the output target. Market clearing therefore does not imply a well-functioning or fully employed economy: it can arise from a rule-imposed ceiling on production. This reflects the post-Keynesian distinction between market clearing and full employment. When firms adjust quantities to expected sales, the goods market can clear at persistently depressed levels of activity, with no endogenous mechanism in the clearing process itself that pushes output or employment toward capacity (Lavoie, 2014).

Unstable monthly adjustment resembles a cost-push process without a stabilizer. With prices tied to realized average cost, a fall in output raises unit costs and prices; the ensuing fall in real purchasing power weakens sales and output further, and the sequence repeats until production has effectively ceased. The post-Keynesian theory of the firm gives this a direct reading: firms are assumed to price on normal-capacity unit costs precisely because pricing on actual unit costs would transmit every fall in utilization into prices (Lavoie, 2014, Ch. 3). The contrast with underemployed market clearing, which differs from this arrangement only in the cost basis, illustrates that argument within the target-limited, monthly-contract setting; it does not extend to other settings. The resulting price–output feedback should not be confused with conflict inflation in the sense of Rowthorn (1977), which arises from competing distributive claims. Here, prices increase mechanically because declining output raises realized unit costs; the model contains neither wage bargaining nor competing wage and profit claims.

Volatile activity is a volatile variant of persistent slack. Like persistent slack, it retains unconstrained production, persistent excess supply, and mild deflation without a cumulative price–output collapse. It differs through a narrower markup band and faster wage and production adjustment. In Kalecki’s markup framework, in which prices are set as a markup over unit costs, the arrangement modifies the markup band and the speed of price and wage adjustment rather than the underlying formation of demand. Over the forty-round horizon, these changes generate greater short-run variation in prices and activity, while excess supply and depressed output persist. The comparison therefore suggests that greater nominal flexibility is insufficient to eliminate a demand-driven shortfall in activity.

Employment does not provide an additional distinction between the arrangements at the present scale: with twelve households, a single contract represents 8.3 percentage points of employment,

leaving the comparison too coarse to identify distinct labor-market adjustment dynamics.

7.2 What language-model households contribute in this design

The language-model layer represents qualitative household differences without assigning a consumption propensity, target balance, or price elasticity to each persona in advance. Under a common prompt context, the selected model produces the spending responses characterized in [Section 5](#). The no-persona reference and the differences between persona means show that the descriptions matter within this specification. The resulting propensities are outputs of the screen, not parameters calibrated to the macro results.

This does not remove behavioral assumptions; it relocates them to the model, persona text, prompt, memory design, and sampling configuration. A rule-based household design could represent comparable heterogeneity with explicit parameters. Whether it would reproduce the same macro paths is not a claim tested here. The relevant advantage is that a household decision can be documented together with its declared rationale.

For the macro runs, the available records link a declared rationale, intended expenditure, realized purchase, and transaction identifier where a purchase occurs, by run, household, and round. They do not retain the rendered prompt or the working-memory snapshot at the time of choice. [Section 5](#) provides one fully documented example of the relationship at the micro level, with its persona text, prompt, and response in [Appendix D](#), but this paper does not use declared rationales to infer a micro-to-macro mechanism. A design intended to analyze complete decision contexts must retain the rendered prompt and contemporaneous memory snapshot for each decision. This would make future shock experiments analyzable at the decision level alongside the aggregate response.

7.3 Scope and limitations

Scale and measurement. The main evaluation contains twelve households and three firms. At this scale, employment is coarse, and under target limitation each firm requires only one worker to meet its production target, causing employment to settle at 25.0% once the initial contracts expire. The resulting measure therefore does not support fine-grained claims about labor-market adjustment across arrangements.

Each arrangement is evaluated over three independent runs. These runs document variation

across simulations but do not support statistical inference, population-level estimates, or model calibration. A supplementary replication with 24 households and six firms reproduces the three qualitative goods-market and price regimes of the primary comparison and preserves the absence of a clear employment separation across the main arrangements (Section F.1). Because the larger block changes population size and persona composition together, it does not isolate a pure scale effect.

The main regimes are characterized over a forty-round horizon. A supplementary 100-round check shows that these forty-round descriptions should not be interpreted as stationary states (Section F.2). Persistent slack develops into a slow deflationary contraction; underemployed market clearing combines a high clearing rate with further declines in output and employment; and unstable monthly adjustment reaches a terminal shutdown. Volatile activity retains substantial real activity over the longer horizon. These extensions qualify the forty-round regimes rather than establish long-run equilibria.

Institutional coverage. The economy has one consumption good, no monetary policy, and no complete government or banking balance-sheet system. The energy constraint is present but nonbinding in the reported macro runs. The evaluated institutional environment is therefore deliberately narrow.

Household and attribution limits. Household decisions depend on the prompt wording, persona text, model version, and sampling settings. The screens characterize behavior within this fixed household specification rather than validating it against observed human behavior. A matched rule-based household comparison would be required to assess substitution between household treatments. The institutional arrangements also change several elements jointly, with the single exception of the price–cost basis within the target-limited pair. The paper therefore reports conditional evidence for complete arrangements, not general individual-rule causality or behavioral fidelity.

8 Conclusion

This paper compares four institutional arrangements while holding one selected language-model household specification fixed in a small synthetic economy. The comparison yields three distinct regimes in prices and goods markets. With unrestricted production, persistent slack combines unsold output with mild deflation. With a production target, goods markets can approach clearing at

a mechanically low level of output and employment. When the same production target and monthly contracts are combined with actual-average-cost rather than normal-capacity pricing, prices and output enter a cumulative collapse in which production effectively ceases in every run. A fourth arrangement that varies markup and adjustment rules under unrestricted production retains the first regime while generating more volatile prices. These are conditional results for complete institutional arrangements, not estimates of a national economy; the one single-rule contrast, between the two cost bases, applies only within the target-limited, monthly-contract setting.

The accompanying micro screens provide an internal characterization of the household treatment. In a common economic context, persona descriptions differentiate spending, and lower liquid balances reduce it. Taken together, the screens provide a contextual, qualitative representation of household differences; they do not validate the responses against observed household behavior.

Within the present model configuration, varying institutional arrangements while holding the characterized household treatment fixed generates persistent slack, low-activity goods-market clearing, and a cumulative cost–price collapse. The contribution is a controlled hybrid comparison that separates the household behavioral channel from the institutional channel. Rule-based macroeconomic ABMs permit institutional experiments but prescribe household decisions *ex ante*; the language-model studies reviewed here make household decisions contextual but treat market, pricing, and contract rules as a fixed backdrop. SIMS combines these elements by holding a contextual, inspectable household treatment constant while varying explicit institutional arrangements. It therefore provides conditional evidence on how complete institutional configurations shape macroeconomic paths, rather than behavioral validation or general individual-rule causality.

A next step is a shock design that records the rendered prompt and contemporaneous memory alongside each household decision. Such an experiment also requires a baseline in which employment, prices, and goods clearing are simultaneously economically plausible. None of the present arrangements provides such a baseline, so they are not used for that purpose. Starting from a more stable baseline, future work could relate changes in individual household decisions to the resulting aggregate response while retaining the complete decision context.

References

- Aher, G., Arriaga, R. I., and Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *PMLR*, pages 337–371.
- Caiani, A., Godin, A., Caverzasi, E., Gallegati, M., Kinsella, S., and Stiglitz, J. E. (2016). Agent based-stock flow consistent macroeconomics: Towards a benchmark model. *Journal of Economic Dynamics and Control*, 69:375–408.
- Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.
- Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly Journal of Economics*, 112(1):1–55.
- Carroll, C. D. (2001). A theory of the consumption function, with and without liquidity constraints. *Journal of Economic Perspectives*, 15(3):23–45.
- Dosi, G., Fagiolo, G., and Roventini, A. (2010). Schumpeter meeting keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics and Control*, 34(9):1748–1767.
- Dwarakanath, K., Vyetenko, S., Tavallali, P., and Balch, T. (2024). ABIDES-Economist: Agent-based simulator of economic systems with learning agents. *arXiv preprint arXiv:2402.09563*.
- Filippas, A., Horton, J. J., and Manning, B. S. (2024). Large language models as simulated economic agents: What can we learn from Homo Silicus? In *Proceedings of the 25th ACM Conference on Economics and Computation (EC '24)*, pages 614–615, New York, NY, USA. Association for Computing Machinery.
- Godley, W. and Lavoie, M. (2007). *Monetary economics: An integrated approach to credit, money, income, production and wealth*. Palgrave Macmillan.
- Gualdi, S., Tarzia, M., Zamponi, F., and Bouchaud, J.-P. (2015). Tipping points in macroeconomic agent-based models. *Journal of Economic Dynamics and Control*, 50:29–61.
- Hein, E. (2023). *Macroeconomics after Kalecki and Keynes: Post-Keynesian foundations*. Edward Elgar.

- Kalecki, M. (1954). *Theory of economic dynamics: An essay on cyclical and long-run changes in capitalist economy*. George Allen and Unwin.
- Lavoie, M. (2014). *Post-Keynesian economics: New foundations*. Edward Elgar.
- Li, N., Gao, C., Li, M., Li, Y., and Liao, Q. (2024). 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)*, pages 15523–15536. Association for Computational Linguistics.
- Li, X., Hao, Y., Hou, J., Huang, J., Wen, Q., et al. (2026). MatrAIx: Simulating the world with 8.3 billion persona agents. *arXiv preprint arXiv:2608.04205*.
- Liu, G., Wang, C., Li, J., Wu, H., and Jiang, C. (2026). Do LLM agents really mimic humans? diagnosing and aligning microeconomic behaviors in macro-ABMs. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 36102–36122. Association for Computational Linguistics.
- MatrAIx (2026). MatrAIx persona 1M. Hugging Face. Data set.
- McGraw, K. O. and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30–46.
- Rowthorn, R. E. (1977). Conflict, inflation and money. *Cambridge Journal of Economics*, 1(3):215–239.
- Seppelcher, P. (2012). Flexibility of wages and macroeconomic instability in an agent-based computational model with endogenous money. *Macroeconomic Dynamics*, 16(S2):284–297.
- Tesfatsion, L. and Judd, K. L., editors (2006). *Handbook of computational economics, Volume 2: Agent-based computational economics*. North-Holland.

A Model parameters

The tables below document the parameter values used in all four macro configurations. [Table A.1](#) gives the values held constant across configurations; [Table A.2](#) gives the configuration-specific values. The correspondence between these economic names and the fields of the archived configuration files is recorded with the evidence bundle, not in the paper.

A.1 Baseline parameters

Table A.1: Baseline parameters

Parameter	Value	Description
<i>Economy and market</i>		
Simulation rounds	40	Rounds per run; round 1 = January 2020
Households	12	Fixed throughout the macro evaluation
Firms	3	Fixed throughout the macro evaluation
Initial household balance	\$500	Same for all twelve households
Initial wage rate	\$2.40 per hour	Nominal hourly wage at round 1
Initial goods price	\$4.10	Nominal unit price of consumption good at round 1
Working hours per week	40	Full-time contract
Wage adjustment speed	0.011	Per-round adjustment toward labor-market clearing
Household income tax rate	20%	Flat rate
Corporate tax rate	25%	Flat rate
Unemployment benefit	\$150	Fixed per round: minimum and maximum bound coincide, so the wage-based rule is inactive
Bank reserve ratio	0.10	ρ in the lending-capacity constraint
Loan-to-revenue ceiling	1.2	μ in the revenue constraint
<i>Firm technology and pricing</i>		
Total factor productivity	1.4	Cobb–Douglas scale parameter A
Labor elasticity	0.6	Cobb–Douglas exponent α
Capital elasticity	0.3	Cobb–Douglas exponent β
Markup lower bound (baseline)	2.4	$\underline{m}$

Continued on next page

Table A.1 continued

Parameter	Value	Description
Markup upper bound (baseline)	3.0	$\bar{m}$
Maximum price step	0.05	Upper bound of the uniform price-adjustment draw
Price-raise threshold	0.25	Unsold-to-sold ratio below which a firm may raise its price
Price-cut threshold	1.00	Unsold-to-sold ratio above which a firm may cut its price
Price stickiness	0.5	Probability that an adjustment is suppressed
Capital per worker	50.0	Initial capital-to-labor ratio
Depreciation rate	5%	Per round
Desired unsold-to-sold ratio	0.5	Target ratio of unsold to sold units within a round
Production adjustment rate λ	0.20	Speed of output target revision
Production target rule	trend feedback	Common to all four configurations (Appendix B)
Signal smoothing weight η	0.40	Weight of the current round in the sales and clearing-rate averages
Target limitation	no	Baseline: unconstrained Cobb–Douglas output
Output-constraint pricing	actual avg. cost	Active only under target limitation (inactive in the baseline)
<i>Energy mechanism</i>		
Maximum energy	100.0	Household energy ceiling
Initial energy	100.0	All households start fully energized
Energy cost per work hour	0.040	Energy drain per hour worked
Base energy cost per round	3.0	Fixed per-round drain regardless of work
Consumption recovery scale	1.25	Recovery is $1.25 \times \ln(1 + \text{units consumed})$
Leisure recovery rate	0.020	Energy restored per leisure hour
Shutdown floor	0.0	Threshold below which labor supply is blocked
<i>Household decision</i>		

Continued on next page

Table A.1 continued

Parameter	Value	Description
Decision model	Gemma 4 31B	Instruction-tuned model with publicly released weights, 31 billion parameters
Decision-model temperature	0.7	Sampling temperature for contract and spending
Decision-model sampling	0.9 / 40 / 1,024	Nucleus probability and top- k cutoff (the sampler’s two truncation rules) and the maximum response length in word pieces
Memory-revision model	Qwen3 14B Q4	Model with publicly released weights, 14 billion parameters, compressed to 4-bit precision
Memory-revision temperature	0.0	Deterministic consolidation of retained insights
Memory capacity	7	Maximum retained recent events and insights
Goal reflection	off	Households form no explicit goals

A.2 Configuration-specific parameter patches

The four configurations share the baseline above. Table A.2 lists the parameters that define an arrangement or differ between arrangements; rows that are identical across the four columns are included so that each column reads as a complete specification.

Table A.2: Configuration-specific parameter patches

Parameter	Persistent slack	Underemployed clearing	Unstable monthly	Volatile activity
Contract duration (months)	4	1	1	4
Target limitation	no	yes	yes	no
Output-constraint pricing	—	normal-capacity cost	actual average cost	—
Markup lower bound	2.4	2.4	2.4	2.1
Markup upper bound	3.0	3.0	3.0	2.4
Wage adjustment speed	0.011	0.011	0.011	0.017
Signal smoothing weight η	0.40	0.40	0.40	0.35
Production adjustment rate λ	0.20	0.20	0.20	0.25
Clearing-rate lower bound $\underline{s}$	0.40	0.40	0.40	0.30
Unsold-share trend damping γ	0.10	0.10	0.10	0.20

B Core model equations

This appendix presents the mathematical form of the four core mechanisms that determine firm-level supply and the market-level adjustment of wages and prices. Parameter values refer to [Table A.1](#) and [Table A.2](#). All equations operate at the monthly round frequency.

B.1 Production technology

[Section 3.4](#) defines potential output Q_t^{pot} with a two-factor Cobb–Douglas technology. When the target-limitation rule is active, realized output is capped at the firm’s rule-based production target $\bar{Q}_t$ from [Section B.4](#):

$$Q_t = \min\left(Q_t^{\text{pot}}, \bar{Q}_t\right). \quad (6)$$

B.2 Wage adjustment

The nominal wage adjusts each round in proportion to labor-market tightness. Tightness is measured as excess demand relative to supply, both expressed in worker-equivalents:

$$\theta_t = \frac{D_t^L - S_t^L}{S_t^L}, \quad (7)$$

where D_t^L is total worker demand across all firms, limited by each firm’s cash, and S_t^L is the total weekly hours supplied by households divided by the 40-hour contract week. The wage then follows a proportional partial-adjustment rule:

$$w_{t+1} = w_t (1 + \delta_w \theta_t), \quad (8)$$

with adjustment speed $\delta_w = 0.011$, or 0.017 under volatile activity. Positive tightness (excess demand) raises the wage; negative tightness (excess supply) lowers it. The result is bounded below at 0.01.

B.3 Price adjustment and cost floor

Firms set prices within a markup band over a unit-cost floor. The cost floor c_t is the wage bill of the round per unit of output:

$$c_t = \begin{cases} \frac{W_t}{Q_t^{\text{pot}}} & \text{normal-capacity-cost pricing while the target binds,} \\ \frac{W_t}{Q_t} & \text{otherwise.} \end{cases} \quad (9)$$

The first case applies to underemployed clearing. Here, W_t is total wage expenditure in round t , Q_t is realized output, and Q_t^{pot} is Cobb–Douglas output at the full hired labor force before the target cap is applied. In the persistent-slack and volatile-activity configurations, the target does not bind, so realized and potential output coincide and the two bases imply the same floor. The markup band applies in every configuration:

$$p_t \in [\underline{m} c_t, \overline{m} c_t]. \quad (10)$$

The baseline bounds are $\underline{m} = 2.4$ and $\overline{m} = 3.0$, or 2.1 and 2.4 under volatile activity, as reported in [Table A.1](#) and [Table A.2](#).

Within this band, prices respond to the unsold-to-sold ratio of the current round,

$$\iota_t = \frac{U_t}{\max\{S_t, 1\}}, \quad (11)$$

where U_t is the firm’s unsold quantity and S_t its sold quantity. A uniform random draw $u_t \sim U[0, \bar{u}]$, with $\bar{u} = 0.05$, scales the price up or down:

$$p_{t+1} = \begin{cases} p_t(1 + u_t) & \text{if } \iota_t < \underline{\iota} \text{ and } m_t < \overline{m}, \\ p_t(1 - u_t) & \text{if } \iota_t > \bar{\iota} \text{ and } m_t > \underline{m}, \\ p_t & \text{otherwise.} \end{cases} \quad (12)$$

The thresholds are $\underline{\iota} = 0.25$ and $\bar{\iota} = 1.0$, and m_t denotes the current markup over the cost floor. With probability $\kappa = 0.5$, an independent draw suppresses the adjustment regardless of the signal,

making prices sticky. The resulting price is then clamped to the markup band, so a change in the cost floor moves the price even in rounds without an adjustment draw. This clamp is the channel through which a rising actual unit cost drives the price under target limitation.

B.4 Production targeting

All four configurations revise the output target with the same trend-feedback rule; they differ only in its parameters (Table A.2). The target has two roles. It determines each firm's demand for additional workers in every configuration by inverting the production function at the current capital stock. It caps realized output only where target limitation is active.

The rule smooths the firm's clearing rate s_t , defined as units sold over units available, and its sales q_t using exponential moving averages:

$$\hat{s}_t = \eta s_t + (1 - \eta) \hat{s}_{t-1}, \quad (13)$$

$$\hat{q}_t = \eta q_t + (1 - \eta) \hat{q}_{t-1}. \quad (14)$$

The smoothing weight is $\eta = 0.40$, or 0.35 under volatile activity. The demand base for the next round depends on whether the smoothed clearing rate meets the target $s^* = 0.50$:

$$B_t = \begin{cases} \max(D_t^{\text{recent}}, \hat{q}_t) & \text{if } \hat{s}_t \geq s^*, \\ \hat{q}_t & \text{if } \hat{s}_t < s^*. \end{cases} \quad (15)$$

Here, D_t^{recent} is the larger of units sold and the firm's share of desired household spending at the current price. Before any sale has occurred, the base is a fixed fraction, 0.40, of full-capacity output. The raw target is

$$\bar{Q}_{t+1} = B_t (1 + \lambda(\iota^* - \iota_t)), \quad (16)$$

with adjustment rate $\lambda = 0.20$, or 0.25 under volatile activity, desired unsold-to-sold ratio $\iota^* = 0.50$, and ι_t the current unsold-to-sold ratio.

Three further adjustments apply once sales history exists. If the smoothed clearing rate is below target, the target is scaled down by

$$\max \left\{ \frac{\hat{s}_t}{s^*}, \underline{s} \right\}, \quad (17)$$

with $\underline{s} = 0.40$, or 0.30 under volatile activity. If the unsold-to-sold ratio has risen relative to the previous round, the target is multiplied by $1 - \gamma$, with $\gamma = 0.10$, or 0.20 under volatile activity. If the smoothed clearing rate is below $\underline{s}$, the target is capped at $\underline{s}$ times full-capacity output. The final target is clamped to

$$0 \leq \bar{Q}_{t+1} \leq Q_t^{\text{cap}}, \quad (18)$$

where Q_t^{cap} is Cobb–Douglas output at the current capital stock with all four worker positions filled.

C Measurement conventions

Employment and goods clearing. Employment is the share of households under an active contract. With twelve households, one contract corresponds to 8.3 percentage points. The goods-clearing rate is the share of available output sold. Because unsold output is not carried over, available output equals production in every round. The late-round goods-clearing rate is calculated as total units sold divided by total units produced over rounds 36–40, rather than as the mean of the five per-round clearing rates.

Prices and output. Mean price growth is the arithmetic mean of the per-round growth rates of the goods price level over rounds 6–40. Inflation volatility is the sample standard deviation of the same series, expressed in percentage points. Smoothed output growth is calculated from real production. For each round from 20 to 40, mean output over the preceding five rounds is compared with mean output over the five rounds before that; the ratio is converted to a per-round compound growth rate, and the resulting twenty-one rates are averaged.

Energy and participation. Paid work and a fixed monthly cost reduce household energy, while leisure and goods consumption restore it. Households below the participation threshold cannot supply labor. Across the twelve reported macro runs, mean household energy averaged over rounds 6–40 lies between 83 and 98 out of 100. The lowest single-round cross-household mean is 75.3. No energy shutdown occurs, so the constraint never excludes a household from labor supply and is non-binding for the macro comparison.

D Micro-screen decision prompt

The micro screens combine a selected persona description with a common decision context. Fields in braces vary by persona or experimental condition. Memory and recent events are empty by design. This appendix documents the micro-screen prompt, not a rendered macro decision prompt, which also includes the household’s simulated state and retained memory. The screen’s January 2026 context is independent of the macro simulation calendar, which begins in January 2020.

System message

You are {agent_name}.

{persona_description}

=== Memory ===

No consolidated experiences yet.

=== Recent Events ===

No recent events.

Decision context

It is January 2026.

=== Financial Position ===

Cash balance: \${balance}

=== Economy Context ===

Goods price: \$4.10/unit

Current wage rate: \$2.40/hour

=== Energy Status ===

Current energy: 100.0.

Consuming goods restores energy after goods market clearing.

How much money do you want to spend on consumption this month?

Money not spent remains in your account.

Consider your profile, past experience, and the information above.

Set decision to a whole-dollar amount from 0 through your cash balance.

Illustrative baseline response

The following is Daniel's first baseline response, rendered from the template above.

Persona

Daniel is a 70-year-old divorced man with adult children. In economic choices, he is cost-sensitive and risk-tolerant. Financial security and enjoyment are moderately important, while wealth is a core value. His prudence and self-regulation are strong, decision style is directive, shopping style is bargain hunter, and attitude toward consumerism is skeptical.

Response

I have \$500 in my account and a current energy level of 100, meaning I do not have an immediate physical need for goods. Given my core value is wealth and my skeptical attitude toward consumerism, I see no reason to spend money unless it is a bargain or a necessity. At \$4.10 per unit, these goods are not a steal. I will keep my money in the account to grow my wealth and maintain my financial security.

Decision: \$0

The record illustrates the decision-level material available from the micro screens; the declared

rationale is not causal evidence.

E Construction of the MatrAIx household personas

This appendix documents the household population used in the micro screens and macro simulations. The profiles are synthetic, not observations of real people, and do not represent a national population.

Source material and retained attributes

The personas are drawn from the decoded 999-row sample distributed with the MatrAIx Persona 1M dataset (MatrAIx, 2026). Only records marked as synthetic are retained; the source-file checksum is recorded in the evidence inventory.

Fields describing a person’s current economic situation are excluded. Income, employment, occupation, housing, financial trajectory, and location would duplicate or conflict with the cash balances, wages, prices, labour opportunities, and other conditions generated by the simulation. The retained attributes describe demographic and preference characteristics:

Table E.1: MatrAIx fields retained in the household persona descriptions

Source field	Use in persona text
Age bracket, gender identity, marital status, children	Brief demographic description
Economic motivation, risk tolerance	General orientation in economic choices
Financial security, wealth, enjoyment values	Stated priorities in financial trade-offs
Prudence, self-regulation	Behavioural dispositions
Decision style, shopping style, consumerism attitude	Decision and consumption orientation

Illustrative profile

The following is Maya’s persona description, used in the micro screens and the twelve-household macro comparison:

Maya is a 21-year-old single woman with two children. In economic choices, she is cost-sensitive and risk-averse. Financial security is a core value; wealth and enjoyment are moderately important. Her prudence is moderate, self-regulation is slight, decision style is consensus-driven, shopping style is impulse buyer, and attitude toward consumerism

is positive.

The profile contains no income, wealth, occupation, location, or numerical spending instruction. These conditions are supplied by the model at each decision. Age brackets are represented by their midpoint; for example, 18–24 becomes 21. This is a schema normalization, not a claim that MatrAIx provides an exact age. First names are readability pseudonyms and are not behavioural inputs.

The twelve-persona population

The household population was selected from the synthetic source sample to cover the available economic motivations and risk attitudes. It is used unchanged in the micro screens and the twelve-household, three-firm macro comparison. The source row of each profile and the population records are listed in the evidence inventory.

This purposeful selection tests whether distinct qualitative profiles generate distinguishable decisions. It is not a population distribution. Appendix F reports a supplementary replication with an independently drawn population of 24 households and six firms. Because that replication changes population size and persona composition together, it is not a pure scale test.

Implication for interpretation

The profiles test whether the selected language model maps qualitative preference attributes to heterogeneous decisions under a common economic context. They neither calibrate the model to an empirical distribution of preferences nor identify causal effects of individual attributes.

F Population and horizon robustness

This appendix reports two supplementary checks of the four institutional arrangements in [Section 6](#). The first changes the population from twelve households and three firms to 24 households and six firms. The second extends the horizon from 40 to 100 rounds. Neither check replaces the primary comparison or is pooled with it.

F.1 Population robustness: 24 households and six firms

The population check repeats the four arrangements with 24 households and six firms, preserving four households per firm, the 40-round horizon, seeds 42–44, the base configuration, the decision prompt, and the household and memory models. The 24 personas are independently drawn from the same synthetic MatrAix source as the twelve-persona population documented in [Appendix E](#), but do not overlap with it.

All twelve arrangement–seed combinations completed. The table compares configuration medians across the three seed-specific runs with the corresponding medians from the twelve-household, three-firm primary comparison.

Table F.1: Population robustness: 24-household, six-firm replication

Arrangement	Mean employment 24/6 · 12/3	Late clearing 24/6 · 12/3	Mean price growth 24/6 · 12/3	Volatility 24/6 · 12/3	Smoothed output growth 24/6 · 12/3
Persistent slack	24.2% · 25.0%	32.2% · 35.1%	−1.23% · −1.21%	0.65 · 0.87	−0.00% · +0.20%
Underemployed market clearing	24.9% · 25.0%	81.9% · 87.6%	−0.83% · −0.81%	0.37 · 0.28	−1.34% · −0.48%
Unstable monthly adjustment	24.8% · 24.5%	n.i. · n.i.	+18.08% · +32.10%	29.50 · 56.81	−14.44% · −18.28%
Volatile activity	27.1% · 25.7%	41.1% · 40.0%	−1.50% · −1.55%	2.60 · 2.54	+0.14% · +0.65%

Notes: Each entry is the median across three seed-specific runs. Mean employment, price growth, and volatility use rounds 6–40; late clearing uses rounds 36–40; smoothed output growth uses rounds 20–40. Volatility is measured in percentage points. n.i. = not interpretable, because production in rounds 36–40 is below one unit per month.

The three qualitative regimes of the primary comparison remain. Persistent slack combines incomplete clearing with mild deflation; underemployed market clearing attains high clearing only at restricted output; and unstable monthly adjustment again produces a price–output collapse. Volatile activity remains a more volatile variant of persistent slack.

The larger population improves employment resolution from 8.3 to 4.2 percentage points per active contract. Nevertheless, mean employment remains close across the three primary arrangements, between 24.2% and 24.9%. The absence of a clear employment separation in the primary

comparison is therefore not removed by the finer resolution. Because the population check changes both scale and persona composition, it is not a pure scale test.

F.2 Horizon robustness: 100 rounds

The horizon check repeats the four arrangements for 100 rounds with the same twelve-household, three-firm population, decision prompt, household and memory models, base configuration, and seeds 42–44. The 100-round runs are independent replications rather than continuations of the runs in Table 3; language-model sampling can therefore produce different early paths despite the shared seeds.

All twelve arrangement–seed combinations completed. The table compares the 100-round configuration medians with the corresponding 40-round medians from the primary comparison. It therefore shows directly which descriptions persist and which change when the observation window is extended.

Table F.2: Horizon robustness: 100-round and 40-round comparison

Arrangement	Mean employment 100 · 40	Late employment 100 · 40	Late clearing 100 · 40	Mean price growth 100 · 40	Volatility 100 · 40	Late production 100 only
Persistent slack	18.3% · 25.0%	10.0% (1.2) · 25.0% (3.0)	45.6% · 35.1%	−0.74% · −1.21%	2.13 · 0.87	180
Underemployed market clearing	19.4% · 25.0%	16.7% (2.0) · 25.0% (3.0)	94.2% · 87.6%	+0.27% · −0.81%	0.89 · 0.28	82
Unstable monthly adjustment	8.3% · 24.5%	8.3% (1.0) · 25.0% (3.0)	n.i. · n.i.	+0.00% · +32.10%	0.00 · 56.81	0
Volatile activity	23.9% · 25.7%	23.3% (2.8) · 25.0% (3.0)	46.2% · 40.0%	−1.45% · −1.55%	3.43 · 2.54	423

Notes: Each entry is the median across three seed-specific runs. For the 100-round comparison, mean employment, price growth, and volatility use rounds 71–100; late employment, clearing, and production use rounds 96–100. The corresponding 40-round measures use the windows reported in Table 3. Parentheses give the late number of employed households. Volatility is measured in percentage points. n.i. = not interpretable, because production is at or below one unit per month in the late window.

The longer horizon qualifies three of the 40-round descriptions. Persistent slack becomes a slow deflationary contraction: employment falls to one or two households and late production falls to 180 units per month. Underemployed market clearing retains high clearing, but this accompanies only two employed households and late production of 82 units per month. Its high clearing rate therefore remains a ratio to shrinking output rather than evidence of adequate demand.

Unstable monthly adjustment reaches terminal shutdown, with production at zero from approximately round 40 onward. Volatile activity alone retains substantial real activity over the longer horizon. It nevertheless differs from persistent slack in several institutional parameters and does

not identify an individual stabilising mechanism.

The two checks show recurrence across three runs per arrangement, not formal inference. They qualify the primary comparison within this particular household treatment and institutional setting; they do not establish long-run equilibria or show how the specific 40-round paths would have continued.