Modeling Semantic Shifts as Economic Shocks: Grounding Large Language Models in DSGE Simulators for Policy Generation and Forecasting

Aditya Dubey*
BITS Pilani, India
f20220231@pilani.bits-pilani.ac.in

Namah Gupta†
BITS Pilani, India
f20220126@pilani.bits-pilani.ac.in

Abstract

While Large Language Models (LLMs) have achieved unprecedented progress in economic narrative analysis, they remain statistical models of language rather than structural models of economic reality. This research investigates the possibility of grounding LLMs—specifically the Qwen-2.5-1.5B architecture—within the structural laws of macroeconomics. We propose a framework where semantic shifts in online discourse are operationalized as exogenous economic shocks. We deploy a Deep Reinforcement Learning (DRL) agent and subject it to three rigorous optimization paradigms: Proximal Policy Optimization (PPO), Group Relative Policy Optimization (GRPO-1), and an advanced, crisis-amplified semantic-action GRPO (GRPO-2). By replacing physical simulators with Dynamic Stochastic General Equilibrium (DSGE) models, we investigate whether LLMs can internalize macroeconomic "physics." Despite using "Sledgehammer" techniques—crisis amplification and semantic action-tagging—we document a consistent structural failure in learning. Through a decisive calibration grid search, we demonstrate that DSGE environments exhibit an "Inversion Paradox," where stabilizing actions are penalized by short-horizon reward functions. This document provides a exhaustive record of our negative results, establishing that the failure is not algorithmic but inherent to the temporal structure of macroeconomic feedback loops.

1 Introduction and Intellectual Motivation

The rapid ascent of Large Language Models (LLMs) has transformed the analysis of economic narratives. LLMs are now deployed to parse central bank communications, financial news, and social media to extract sentiment and identify latent macroeconomic themes. However, these models essentially operate as advanced correlation engines; they do not internalize the structural laws that govern economic systems—equilibrium conditions, intertemporal budget constraints, or the nonlinear propagation of shocks.

1.1 The Grounding Gap in Economic AI

A fundamental limitation in the current paradigm is that LLMs are trained to model distributions over text, not distributions over economic outcomes. They learn correlations between words and concepts, but they do not learn the structural laws that define stability. As a result, an LLM can generate policy recommendations that are linguistically plausible yet economically incoherent when evaluated inside a formal macroeconomic model. This observation motivates our core question: **Can a language model be grounded in economic structure rather than merely economic language?**

*Equal contribution.
†Equal contribution.

Preprint. Preliminary work.

1.2 The Robotics Analogy: Embodied Economic AI

Our inspiration comes from robotics and embodied AI. In robotics, agents do not learn physics from text descriptions of Newton's laws; they interact with a physics simulator, take actions, observe consequences, and gradually internalize the physical world's structure. Recent work such as GLAM (Grounding Large Language Models in Interactive Environments with Online Reinforcement Learning) shows that similar grounding is possible for LLMs when they are placed inside interactive environments and trained via RL.

This project adopts the same philosophy but replaces the physics simulator with a macroeconomic simulator (DSGE), and replaces physical actions (torques) with policy actions (shocks). Semantic shifts in economic discourse are treated as exogenous shocks, and the LLM is trained to generate policy interventions whose quality is evaluated by their macroeconomic consequences. The central goal is to investigate whether an LLM can learn the internal "physics" of a DSGE model through interaction.

2 Related Work and Conceptual Background

This research is situated at the intersection of Financial NLP, Macroeconomic Theory, and Deep Reinforcement Learning. We review eight key works that inform our methodology.

1. Grounding LLMs in Interactive Environments (GLAM) [1]: GLAM provides the methodological template for grounding. It proves that online RL can align language representations with environment dynamics. We adapt their pipeline—mapping language signals to policy generation—within the "stiff" environment of a DSGE simulator.

2. Detecting Subtle Semantic Shifts [5]: This research (NAACL 2024) focuses on detecting conceptual drifts in financial narratives. It serves as the foundation for our "Semantic Shift Watchdog" module, which quantifies Temporal Semantic Volatility (TSV). Unlike their correlational approach, we operationalize these shifts as inputs to an interactive system.

3. Macroeconomic Narratives from Social Media [4]: This work validates the extraction of economic signals from platforms like Twitter/X and Reddit. We utilize this to justify our use of online discourse as a proxy for exogenous shock identifiers.

4. Causal Exploration of Market Shocks [2]: This work delves into the causal relationship between market shocks and semantic shifts across partisan groups. Understanding these causal links is vital for our project's goal of treating semantic shifts as exogenous shocks rather than just market sentiment.

5. Quantifying Semantic Shift in Financial NLP [6]: This paper addresses the quantification of shifts using Transformer models and DistilBERT. The robust metrics discussed for market prediction stability are instrumental in developing our TSV methodology and ensuring the reliability of our "semantic shock" input vectors.

6. Large Legislative Models [3]: This research aligns with our DRL Policy Agent, exploring how AI can generate efficient policies within economic simulations. It provides a valuable starting point for mapping legislative intents to structural simulator parameters.

7. Deep RL in a Monetary Model [boe2025drl]: The Bank of England demonstrated that RL can solve DSGE-style monetary models using numeric neural networks. Our work differs by utilizing a full-scale LLM as the controller, bridging the gap between language and structural shocks.

8. Identifying Monetary Policy Shocks [nber2024shocks]: This natural language approach identifies policy shocks from FOMC documents. They build NLP-derived shock measures and link them to economic outcomes. We extend this by embedding these signals into an interactive simulator for real-time policy generation.

3 Problem I: Building the RL-Compatible DSGE Environment

3.1 Infrastructural Motivation

Reinforcement learning requires an environment that supports repeated interaction, dynamic interventions, and counterfactual rollouts. Most DSGE tools (like Dynare) are designed for offline analysis.

The first problem was infrastructural: how to turn a DSGE model into an interactive "Gym-like" environment.

3.2 The GenericDSGESimulator Wrapper

We built a wrapper around **Snowdrop** [**snowdrop2025**] with four conceptual responsibilities:

1. **Shock Abstraction:** The simulator accepts shocks as scalars, paths, or sparse (time, value) tuples. This is critical for modeling multi-period policy interventions.
2. **Parameter Updates:** Support for structural updates (e.g., Taylor rule coefficients) enables learning over regimes.
3. **State Cloning:** Deep cloning of calibrations allow multiple counterfactual trajectories to be evaluated from the same initial state, essential for GRPO.
4. **RL Step Interface:** A transition interface that enables integration with RL loops while preserving the DSGE temporal structure.

3.3 Simulator Validation Experiments

A comprehensive validation suite was executed across linear (QPM) and nonlinear (Smets-Wouters) models. The outputs (see Listing 1) show that large shocks produce strong responses followed by gradual mean reversion.

```
=== START TEST SUITE ===
TEST 1 - Nominal Interest Rate Shock (SHK_RS: 1.0)
            RS          RR   DLA_CPI
1998-01-01  16.010000   8.461200  7.250700
1998-04-01  13.272388  11.004444  5.525716
...
TEST 10 - Nonlinear SW model (Multiple Shocks)
            y       pinf         r
2000-01-01  0.000000  0.000000  0.000000
2000-04-01 -0.035428 -0.007073  0.020122
```

Listing 1: Simulator Verification Logs

Conclusion of Phase I: The infrastructure works correctly. However, near equilibrium, the simulator suppresses marginal effects of small actions, implying that reward gradients are near zero. This "physics" of the system sets up the failure of standard RL methods.

4 Problem II: PPO — Formalism and Failure

4.1 Mathematical Formalism

Proximal Policy Optimization (PPO) optimizes the clipped surrogate objective:

$$\mathcal{L}_{PPO}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta) A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A_t \right) \right] \tag{1}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ and A_t is the advantage estimated by a critic.

4.2 The RL Loop Design

In our setup, the State $\mathbf{s}$ consists of macroeconomic variables (inflation, output gap). The Action $\mathbf{a}$ is a structured policy shock. The Transition is the DSGE simulation forward in time. The Reward R is the negative weighted sum of squared deviations from targets.

4.3 Experimental Failure

PPO consistently collapsed to inaction. Generated actions shrank toward zero, variance vanished, and the critic produced unstable value estimates. This is due to:

- Near equilibrium, the advantage A_t is approximately zero.
- Delayed macroeconomic effects (e.g., policy transmission lag) break credit assignment.
- The critic amplifies noise in the "stiff" reward landscape of the DSGE model.

5 Problem III: GRPO-1 (The First Relative Attempt)

5.1 Mathematical Formalism

Group Relative Policy Optimization (GRPO) removes the critic and computes relative advantages:

$$\hat{A}_i = \frac{R_i - \mu_R}{\sigma_R} \tag{2}$$

where R_i is the reward for the i-th sample in a group of size G.

5.2 Why GRPO-1?

Removing the critic resolves the instability issues seen in PPO. By comparing $G = 6$ completions against each other, the model should ideally learn which "direction" of interest rate change leads to lower cumulative loss.

5.3 Observed Failure

GRPO-1 showed higher reward variance than PPO, but learning failed to accumulate. Improvements in one epoch (e.g., Epoch 2) were lost in the next. The failure suggested that the optimization was mechanically correct but the environment was providing insufficient signal.

6 Problem IV: GRPO-2 — Crisis-Amplified Grounding

6.1 Motivation: Disambiguating the Failure

Two hypotheses remained: (1) *Equilibrium Hypothesis*: actions have no effect near steady state. (2) *Noise Hypothesis*: the system is too sensitive. GRPO-2 was designed to forcefully break equilibrium and amplify signal.

6.2 Core Design Principles of GRPO-2

1. **Semantic Action Abstraction:** The LLM emits descriptors: Policy (MONETARY), Direction (TIGHTEN/EASE), and Intensity (MILD to EMERGENCY). These are deterministically mapped to shocks, reducing action noise.
2. **Crisis-Amplified Environment:** We introduced hyperinflation and depression scenarios. Episodes start with a "burn-in" where the crisis unfolds without intervention, ensuring the system is far from equilibrium.
3. **Persistent Actions:** Actions are no longer single impulses; they are injected over multiple quarters with slow decay ($\rho \approx 0.9$).
4. **Inaction Penalty:** A fixed negative reward is applied if the agent produces no action during high-loss states.

6.3 Training Logs and Failure

Logs show intermittent improvements but frequent parse failures and reversion to inaction. This suggests the model is "risk-averse"—it learns that even a "correct" action often worsens the loss function in the short-run, favoring the inaction penalty over the risk of destabilization.

7 Results: The Calibration Grid Search

The calibration grid search removed the LLM and manually probed the environment. This is the project's decisive experiment.

Table 1: Calibration Grid Search: Stabilization vs. Loss Function.

Scenario	Magnitude	Weighting	Base Loss	Act Loss	Delta	Valid?
Recession (Cut)	0.05	STANDARD	216	219	-3.0	FAIL
	0.50	STANDARD	216	248	-32.0	FAIL
Stagflation (Hike)	0.05	HAWK	334	333	1.6	PASS
	0.50	STANDARD	369	378	-9.6	FAIL
Cost-Push (Hike)	0.25	STANDARD	2664	2996	-331.8	FAIL

7.1 Row-Level Interpretation: The Inversion Paradox

The grid search (Table 1) reveals a structural mismatch.

- **Recession:** Economically intuitive cuts *increase* loss because expansionary policy worsens inflation/volatility before stabilizing output. Short-horizon loss functions cannot "see" the eventual recovery.
- **Stagflation:** Policy hikes only help if the reward function is extremely skewed toward inflation (HAWK). Under standard weights, the gradient flips sign depending on the relative weight of GDP vs Inflation.
- **Cost-Push:** No action improves the loss. The DSGE model's short-run dynamics fundamentally punish intervention under this shock.

8 Final Conclusion and Structural Implications

This work provides a rigorous diagnosis of the failure of grounding LLMs in DSGE models via RL. The failure is not algorithmic; it is structural. DSGE models are built around equilibrium restoration, delayed benefits, and short-run pain for long-run stability. Policy-gradient RL, conversely, relies on locally informative rewards and smooth improvement landscapes. These two frameworks are mathematically incompatible. Future work must move toward **Model-Based Planning** (tree search) or **Inverse Optimal Control** to achieve economic grounding.

References

[1] Thomas Carta et al. "Grounding Large Language Models in Interactive Environments with Online Reinforcement Learning (GLAM)". In: *Proceedings of the International Conference on Machine Learning*. 2023. URL: https://proceedings.mlr.press/v202/carta23a/carta23a.pdf. 2

[2] Felix Drinkall et al. "Stories that (are) Move(d by) Markets: A Causal Exploration of Market Shocks and Semantic Shifts across Different Partisan Groups". In: *ACL Findings* (2025). URL: https://arxiv.org/pdf/2502.14497. 2

[3] Henry Gasztowtt et al. "Large Legislative Models: Towards Efficient AI Policymaking in Economic Simulations". In: *arXiv preprint arXiv:2410.08345* (2024). URL: https://arxiv.org/pdf/2410.08345. 2

[4] Almog Gueta et al. "Can LLMs Learn Macroeconomic Narratives from Social Media?" In: *NAACL Findings* (2024). URL: https://arxiv.org/pdf/2406.12109. 2

[5] Jiaxin Liu, Yi Yang, and Kar Yan Tam. "Beyond Surface Similarity: Detecting Subtle Semantic Shifts in Financial Narratives". In: *NAACL Findings* (2024). URL: https://arxiv.org/abs/2403.14341. 2

[6] Zhongtian Sun et al. "Quantifying Semantic Shift in Financial NLP: Robust Metrics for Market Prediction Stability". In: *ACM ICAIF* (2025). URL: https://arxiv.org/pdf/2510.00205. 2