Afriat Sets vs Narrative Rules

Counterfactual Prediction in a Portfolio Choice Experiment
(Paper + Dashboard "Blaper")

Abstract

This paper studies counterfactual prediction in a repeated portfolio-allocation environment where each subject allocates a fresh \$100 across three state-contingent assets under changing price vectors. I compare two paradigms: (i) Afriat/Varian revealed preference, which delivers a concave-utility *correspondence* of counterfactual optimizers, and (ii) a narrative/procedural approach that predicts actions via simple, interpretable rules learned from the subject's past decisions. To evaluate Afriat fairly, I treat it as set-valued and evaluate its counterfactual correspondence as a *set*: for each asset I project the correspondence onto that asset's dollar spend and compute an Afriat band plus a distance-to-band metric. Across 154 subjects (training on trials 1–25; evaluation at trial 50), these projected Afriat sets are often sharp (median widths around \$4–\$5) but frequently exclude realized behavior (coverage rates around 19–23% depending on the asset). On point prediction (train 1–49, predict 50), narrative rule inference materially outperforms Afriat variants in mean L^1 error.

Main synthesis. *Afriat is a constraint; narratives are a selector.* Afriat provides disciplined outer structure for counterfactuals (a feasible set). Narrative procedures provide a selection principle and capture heterogeneity in how subjects behave. In these data, the main failure mode is often not tie-breaking inside a correct Afriat set, but the Afriat set missing the realized action.

Contents

1 Reading guide and a translation table

This paper is intentionally bilingual.

If you come from revealed preference: start with Section **??** (the prediction games) and Section **??** (narrative rules as model selection). Point prediction at $t = 50$ is where Afriat is forced to choose a point (Section **??**).

If you come from LLMs / AI: start with Section **??** for what Afriat is doing (and why it produces sets), then Section **??** for the set-diagnostics dashboard.

Table 1: A Rosetta stone: revealed preference vs prompting language

Econometrics / RP	Prompting / ML analogue	What we do here
Observation (trial) t	One row in the prompt "history"	`t=.. prices ...	spend ...`
Covariates (prices p_t)	Context tokens	price vector shown each period
Action (bundle x_t)	Model output (JSON)	dollar allocation across three "stocks"
Counterfactual query price p_0	Test-time query	reveal true prices at $t = 50$; ask for allocation
Set / partial identification	Non-uniqueness	Afriat gives a correspondence; we score via bands
Model selection	Program selection	fixed rule library; personalize via cross-validation

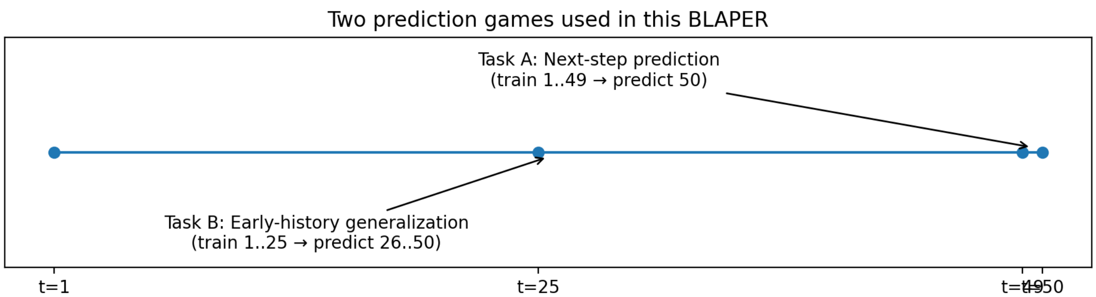


Figure 1: Two prediction games used in the BLAPER. Task A: train 1–49, predict 50 (one-step). Task B: train 1–25, generalize to later price regimes (26–50); the set-diagnostics focus on $t = 50$.

Glossary (quick definitions)

GARP
Generalized Axiom of Revealed Preference: no cycles in the revealed-preference relation (a nonparametric test of concave utility rationalizability).

Afriat inequalities
Linear inequalities whose feasibility is equivalent to GARP; a feasible solution constructs a concave utility that rationalizes the data.

Demand correspondence
A set of optimal bundles at a given price vector. For piecewise-linear utilities, the correspondence is typically set-valued.

Slack δ or δ_t
Additive relaxations that ensure feasibility when exact GARP fails. δ_t is observation-specific.

Afriat band
For a given counterfactual price vector and a given asset, the interval of feasible counterfactual spends implied by Afriat (after projecting the correspondence onto that coordinate).

Coverage
Fraction of subjects whose realized counterfactual spend lies inside the Afriat band.

Distance-to-band
If the truth lies outside the Afriat band, the distance (in dollars) to the nearest band endpoint; 0 if covered.

Narrative rule
A compact subject-specific procedure (e.g., equal-dollar, kNN memory, power-share) inferred from that subject's past choices and then applied to the counterfactual prices.

L^1 error
For point prediction, the sum of absolute dollar errors across the three spends at $t = 50$.

2 Data provenance and experimental design

2.1 Origin of the experiment

The data come from the portfolio-choice experiment of Ahn, Choi, Gale, and Kariv (2014).[**?**] Subjects repeatedly choose portfolios over three Arrow securities (three states of nature). State 2 has known probability $\pi_2 = 1/3$, while states 1 and 3 have unknown probabilities $\pi_1, \pi_3 \geq 0$ with $\pi_1 + \pi_3 = 2/3$.[**?**, p. 196] In each decision problem the subject chooses a nonnegative portfolio $x = (x_1, x_2, x_3) \geq 0$ on a linear budget set $p \cdot x = 1$ (Ahn et al.'s normalization).[**?**, p. 196]

2.2 Experimental procedure (what matters here)

Ahn et al. report:

- 154 subjects at UC Berkeley; each session consisted of **50 independent decision problems**.[**?**, p. 200]
- Each budget set was generated with **intercepts between 0 and 100 tokens**, and **at least one intercept greater than 50 tokens**.[**?**, p. 200]
- Choices were restricted to the budget plane and recorded via a graphical "point-and-click" interface.[**?**, p. 200]
- Payoffs: one decision round was selected at random for payment.[**?**, p. 201]

2.3 Clean file used in this paper

This paper uses the cleaned dataset `rationalitydata3goods.csv` with 7700 rows (154 subjects × 50 trials). Variables:

- `id`: subject identifier; `obs`: trial index $t \in \{1, \ldots, 50\}$.
- `x,y,z`: chosen quantities (x_{Xt}, x_{Yt}, x_{Zt}).
- `xa,ya,za`: budget intercepts (max feasible quantity of each good if all budget is allocated to that good).
- `px,py,pz`: prices, with `px` $\approx 100/$`xa` etc.

Empirically, the intercept constraints match the protocol: across goods, intercepts lie in $[10, 100]$ and each trial has at least one intercept ≥ 50. Because choices are recorded on a discrete interface, budget feasibility holds up to small rounding slack (median slack ≈ 0.28 dollars on a \$100 budget in this cleaned file).

2.4 Narrative re-framing used for prompts

For prediction we convert each trial into dollar spends:

$$s_{it} = p_{it} x_{it}, \qquad i \in \{X, Y, Z\},$$

and define leftover cash $c_t = 100 - (s_{Xt} + s_{Yt} + s_{Zt})$. We re-label goods as "stocks" purely to encourage general reasoning in language models:

$$X = \textsc{PapayaTech}, \quad Y = \textsc{AxolotlWorks}, \quad Z = \textsc{QuipuQuantum}.$$

This is a change of labels/units only.

3 What is being predicted?

Each subject faces trials $t = 1, \ldots, 49$ with prices $p_t = (p_{Xt}, p_{Yt}, p_{Zt})$ and allocates a fresh budget \$100 across the three stocks, yielding dollar spends (s_{Xt}, s_{Yt}, s_{Zt}) and leftover cash. A new price vector arrives at $t = 50$; the goal is counterfactual prediction.

We use two complementary evaluation tracks:

1. **Set-based Afriat diagnostics (train 1–25, evaluate at 50).** Fit Afriat with observation-specific slack δ_t on early history. Compute the counterfactual demand correspondence at $t = 50$ and evaluate it as a set (bands, coverage, distance-to-band).
2. **Point prediction benchmark (train 1–49, predict 50).** Predict the full $t = 50$ allocation and evaluate with L^1 error (sum of absolute errors across the three spends).

4 Methods

4.1 Mapping dollars to quantities (cash as a fourth good)

Afriat theory is stated in quantities. We map dollars to quantities and add cash as a fourth good with price 1:

$$q_{it} = \frac{s_{it}}{p_{it}}, \qquad x_t = (q_{Xt}, q_{Yt}, q_{Zt}, c_t), \qquad \tilde{p}_t = (p_{Xt}, p_{Yt}, p_{Zt}, 1),$$

so $\tilde{p}_t \cdot x_t = 100$.

4.2 Afriat with observation-specific slack δ_t

To accommodate violations, we use the relaxation:

$$u_s \leq u_t + \lambda_t \, \tilde{p}_t \cdot (x_s - x_t) + \delta_t, \qquad \delta_t \geq 0,$$

estimated by minimizing $\sum_t \delta_t$ subject to these inequalities (an LP). Given (u, λ, δ), define the concave envelope

$$U(x) = \min_t \{u_t + \delta_t + \lambda_t \, \tilde{p}_t \cdot (x - x_t)\}.$$

Table 2: Afriat as a set at $t = 50$ (trained on $t = 1..25$): stock-by-stock diagnostics

Stock	Coverage	Med. width	Med. dist	Mean dist|outside	Narrative med. AE	Narrative bet
PapayaTech	23.4%	4.70	11.14	22.27	11.48	
AxolotlWorks	20.1%	4.42	9.53	20.60	10.57	
QuipuQuantum	18.8%	4.39	8.02	18.57	9.49	

At new prices $\tilde{p}_0$, the counterfactual correspondence is

$$D(\tilde{p}_0) = \arg \max_{x \geq 0,\, \tilde{p}_0 \cdot x \leq 100} U(x),$$

which is typically set-valued.

4.3 Set evaluation via projected bands

Rather than forcing Afriat into a single point, we evaluate it as a set. For each stock $i \in \{X, Y, Z\}$, we project $D(\tilde{p}_{50})$ onto the spend coordinate and compute the band $[\min S_i, \max S_i]$ under near-optimality. We report: (i) coverage, (ii) band width, (iii) distance-to-band.

4.4 Narratives as personalized procedures

A narrative model is a compact subject-specific procedure mapping contexts (prices) to actions (allocations). We restrict to a small interpretable *rule library* and select the best rule per subject via cross-validation on that subject's history.

Rule library (explicit definitions). Let prices be (p_X, p_Y, p_Z) and the per-trial budget be 100. We consider:

- **Equal-dollar:** $s_X = s_Y = s_Z = 100/3$.
- **Equal-share:** equal quantities; in dollars $s_i = 100 \cdot p_i/(p_X + p_Y + p_Z)$.
- **Power-share:** $s_i = 100 \cdot w_i$ with $w_i \propto p_i^{-\alpha}$ (price sensitivity controlled by α; $\alpha = 0$ yields equal-dollar).
- **kNN memory:** predict as an (optionally weighted) average of historical allocations at the K nearest historical price vectors.
- **Dominance (all-in):** allocate (nearly) all budget to one stock (e.g., the cheapest).

The goal is not to maximize flexible fit; it is to produce a *procedure you can read* and compare to Afriat's disciplined feasibility set.

5 Results

5.1 Afriat set diagnostics by stock

Figure **??**–**??** show the distribution of distance-to-band and band width, and how these diagnostics vary with counterfactual price dispersion (ratio of max to min price at $t = 50$).

Table 3: Narrative heterogeneity: families inferred from early history ($t = 1..25$)

Narrative family (selected from $t = 1..25$)	Count	Share
Power-share (smooth)	61	39.6%
Case-based (kNN memory)	41	26.6%
Equal-dollar	34	22.1%
Equal-share	13	8.4%
Dominance (all-in)	5	3.2%

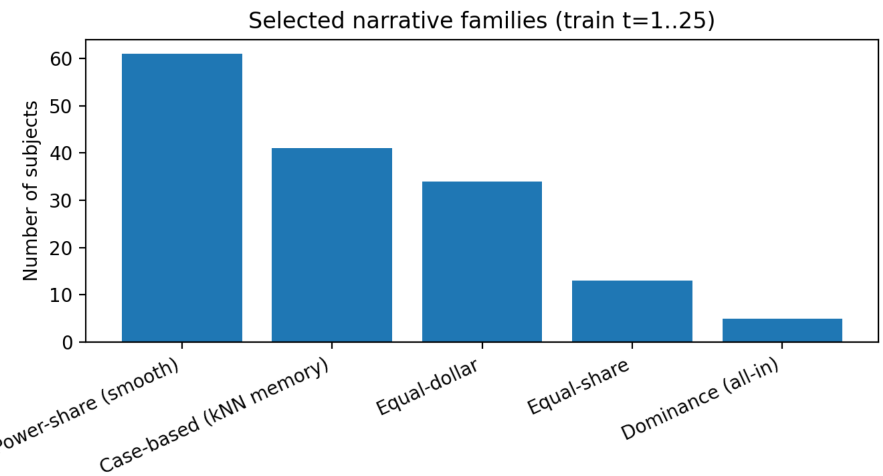


Figure 2: Narrative family counts (selected using early history $t = 1..25$).

5.2 Point prediction benchmark at $t = 50$

5.3 One-subject deep dive (Subject 920)

Figure ?? illustrates a representative "set miss" case: the Afriat band for PapayaTech can be narrow yet far from the truth.

6 The dashboard companion (standalone HTML)

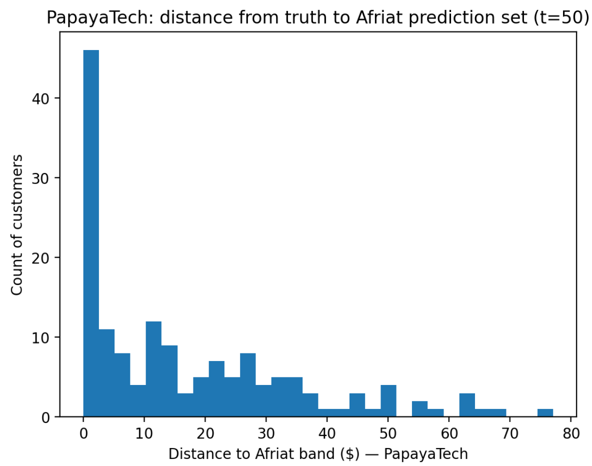
This "blaper" is shipped in two synchronized formats:

- **PDF (this document):** traditional academic narrative + figures/tables.
- **Standalone HTML dashboard:** the same content plus interactive browsing (stock selector, metric hover-glossary, and a subject explorer).

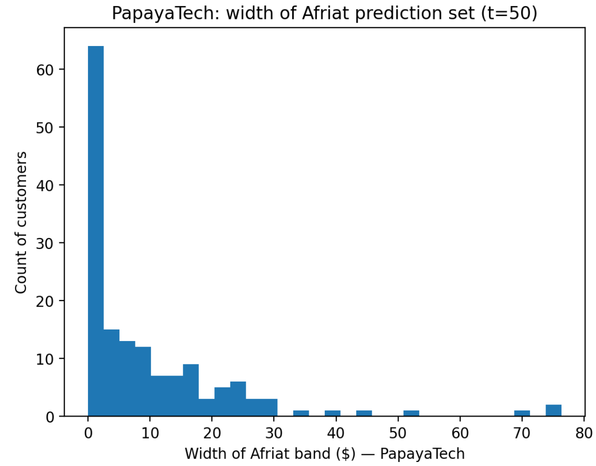
Open `afriat_vs_narrative_BLAPER.html` to use the dashboard view offline.

7 Conclusion

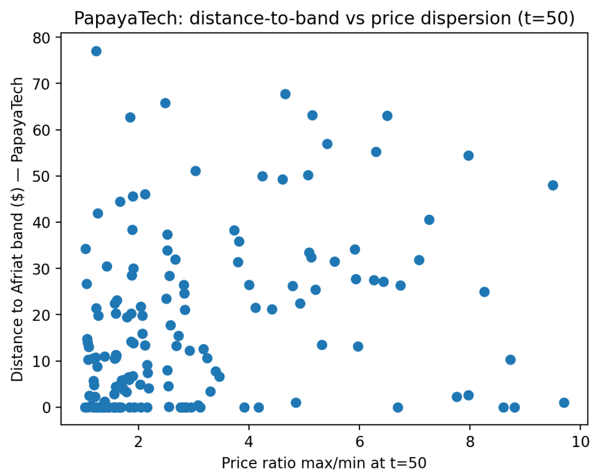
Afriat delivers disciplined outer structure for counterfactuals, but counterfactual prediction is inherently a selection problem and, in procedural environments, often calls for procedural abstractions.



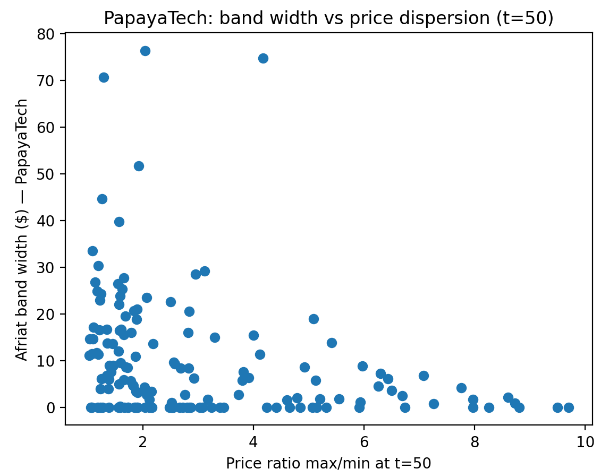
(a) Distance-to-band (t=50)



(b) Band width (t=50)



(c) Distance vs price dispersion



(d) Width vs price dispersion

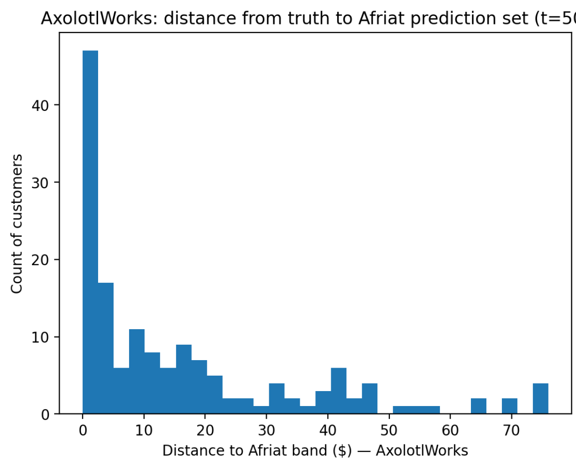
Figure 3: PAPAYATECH (PapayaTech / X): Afriat set diagnostics at $t = 50$ (trained on $t = 1..25$).

Narrative rules provide interpretable, subject-specific procedures that predict well and help diagnose when the Afriat correspondence is missing the realized action.

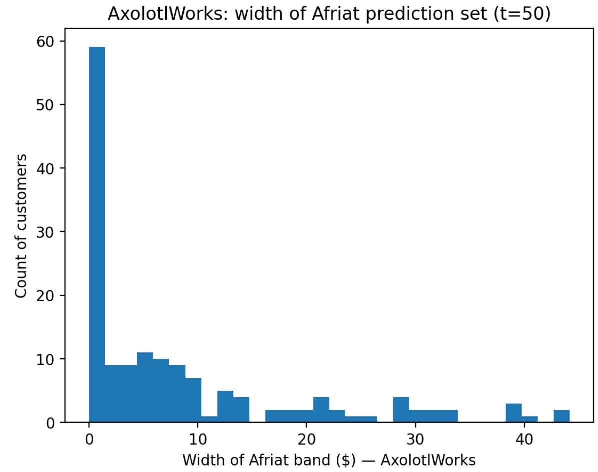
A Narrative prompt template (abridged)

The evaluation prompts follow a stable template. The full prompts include 49 history lines (t=1..49) and ask for a strict JSON prediction at t=50.

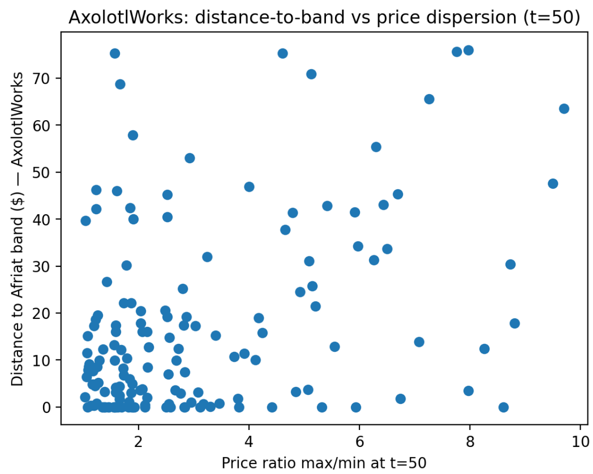
```
Investor Journey - Subject {ID}

Stocks:
  - PapayaTech (ticker: X)
  - AxolotlWorks (ticker: Y)
  - QuipuQuantum (ticker: Z)

Your past 49 trials (t = 1..49):
```



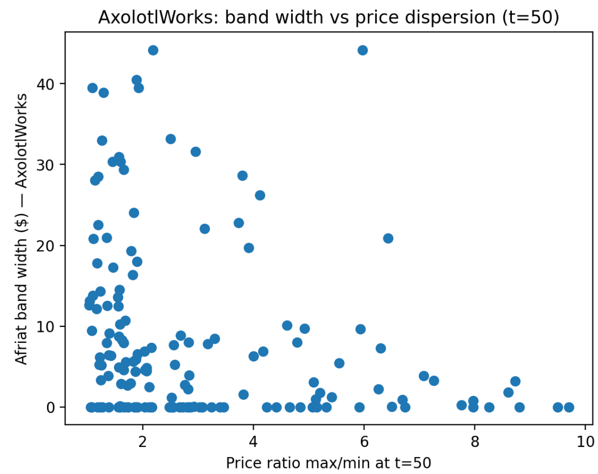
(a) Distance-to-band (t=50)



(b) Band width (t=50)



(c) Distance vs price dispersion



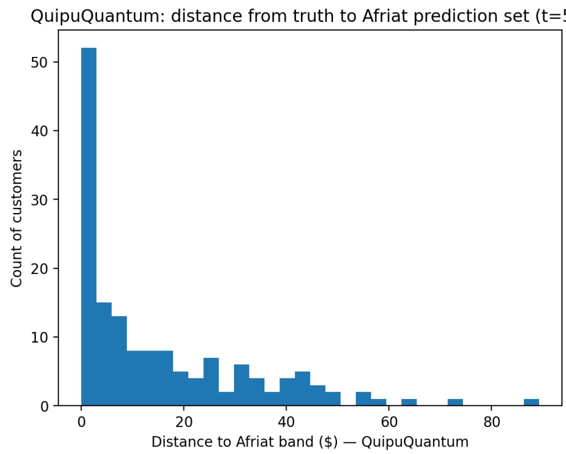
(d) Width vs price dispersion

Figure 4: AXOLOTLWORKS (AxolotlWorks / Y): Afriat set diagnostics at $t = 50$ (trained on $t = 1..25$).

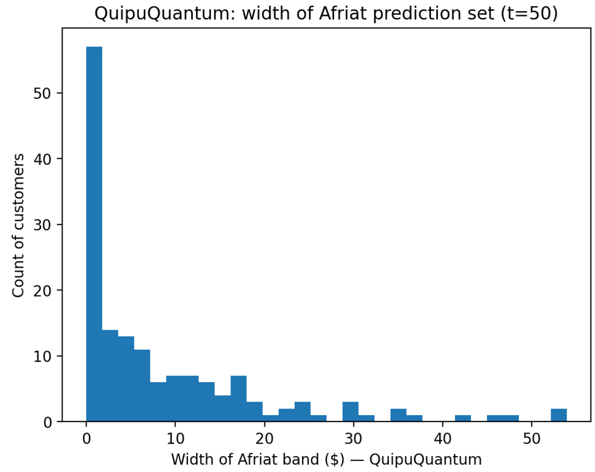
```
t=01 prices: X=$..., Y=$..., Z=$... | spend: X=$..., Y=$..., Z=$... | cash=$...
...
t=49 prices: X=$..., Y=$..., Z=$... | spend: X=$..., Y=$..., Z=$... | cash=$...

Now the 50th trial arrives.
t=50 prices: X=$pX50, Y=$pY50, Z=$pZ50

Return ONLY a single JSON object:
{"PapayaTech": 0.00, "AxolotlWorks": 0.00, "QuipuQuantum": 0.00, "Cash": 0.00}
```

References

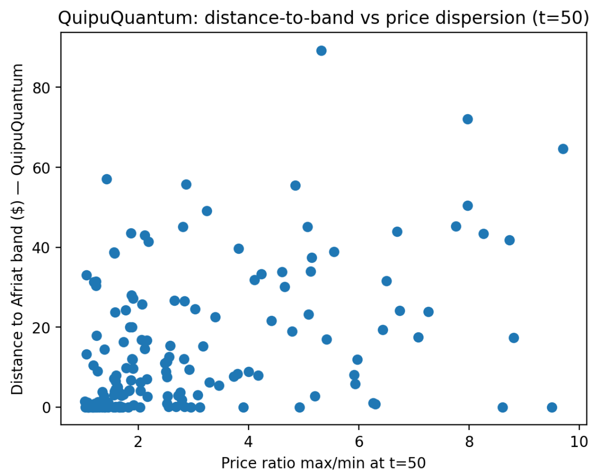
[1] David Ahn, Syngjoo Choi, Douglas Gale, and Shachar Kariv. Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics*, 5(2):195–223, 2014.



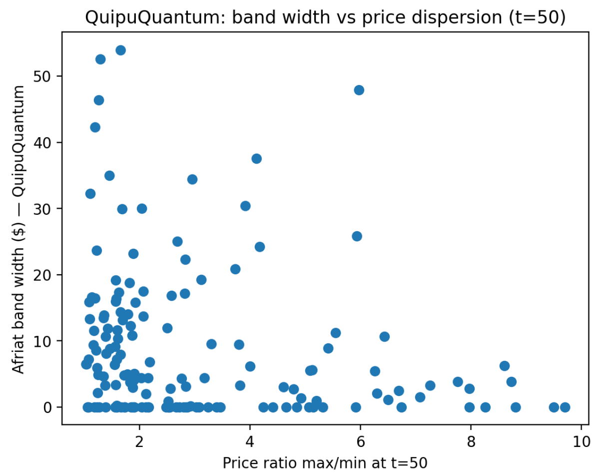
(a) Distance-to-band (t=50)



(b) Band width (t=50)



(c) Distance vs price dispersion



(d) Width vs price dispersion

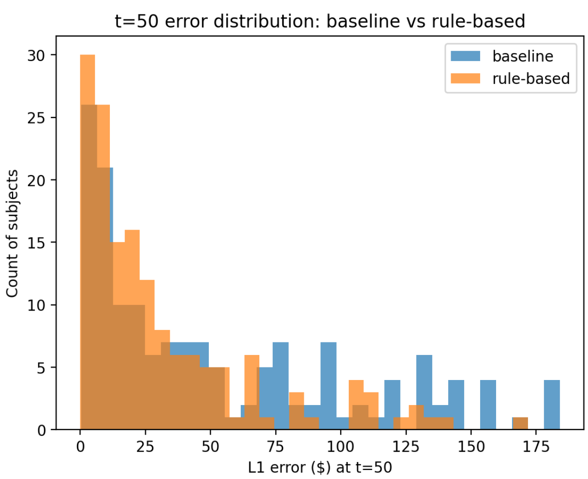
Figure 5: QuipuQuantum (QuipuQuantum / Z): Afriat set diagnostics at $t = 50$ (trained on $t = 1..25$).

[2] S. N. Afriat. The construction of a utility function from expenditure data. *International Economic Review*, 8(1):67–77, 1967.

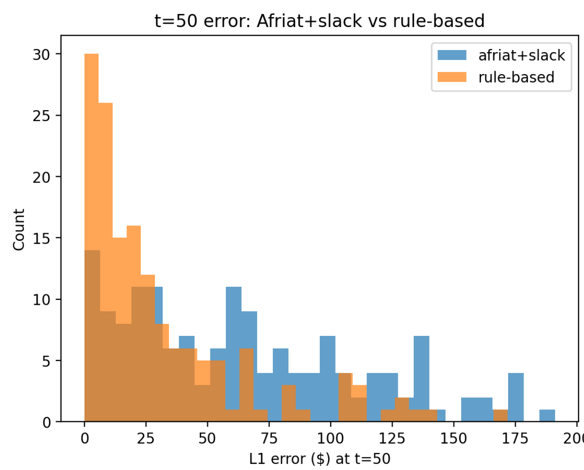
[3] H. R. Varian. The nonparametric approach to demand analysis. *Econometrica*, 50(4):945–973, 1982.

Table 4: Point prediction at $t = 50$: narrative rules vs Afriat variants (154 subjects)

Model	Mean L^1	Median L^1	90th pct L^1	Top-asset acc.	Near all-in pred.	Tr
Rule-based (CV narrative rules)	31.34	18.50	83.22	55.2%	4.5%	
Afriat per-t δ_t + tie-break (closest to anchor) + 0.1-share grid	47.80	31.82	115.88	53.9%	9.7%	
Baseline (mean-share scaling)	52.51	33.82	134.34	44.2%	0.0%	
Afriat per-t δ_t + 0.1-share postprocess	52.66	43.67	111.56	49.4%	13.0%	
Afriat global δ + 0.1-share postprocess	58.70	53.43	121.65	45.5%	18.2%	
Afriat + per-t δ_t (additive)	58.76	50.73	127.07	52.6%	8.4%	
Afriat + global δ (additive)	64.15	58.72	134.62	44.2%	14.3%	



(a) Baseline vs narrative: L^1 error



(b) Afriat vs narrative: L^1 error

Figure 6: Point prediction at $t = 50$: distributions of L^1 error.

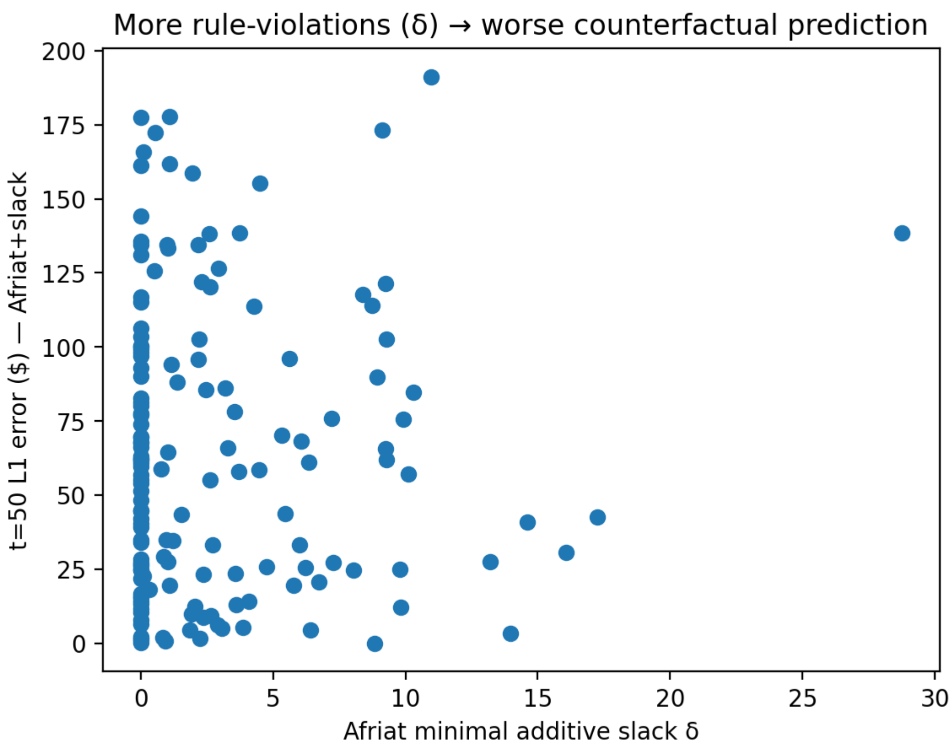


Figure 7: L^1 error vs fitted Afriat slack at $t = 50$: larger fitted slack correlates with worse counterfactual prediction.

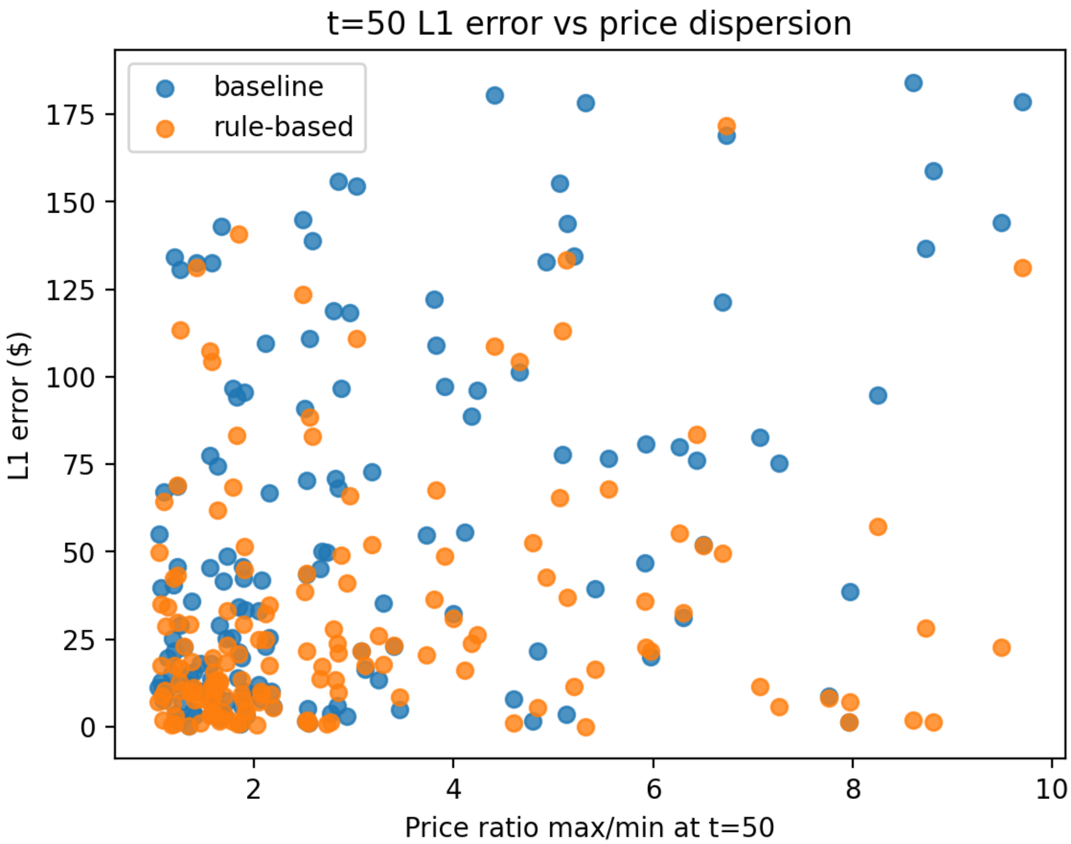
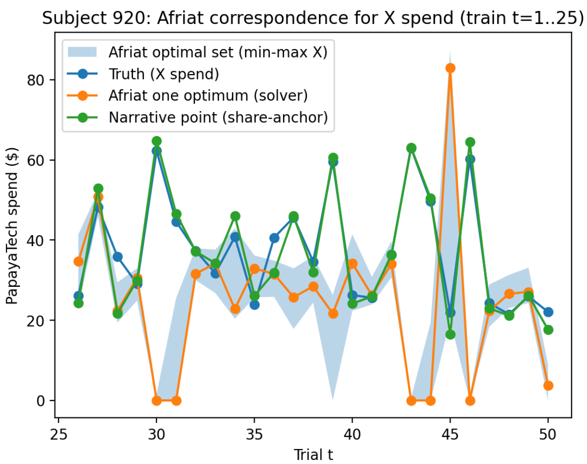
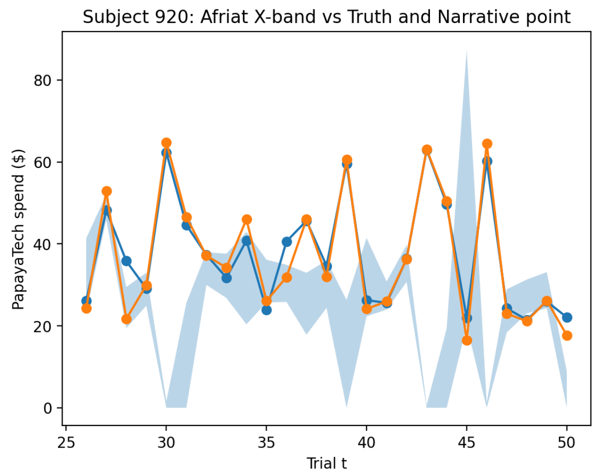


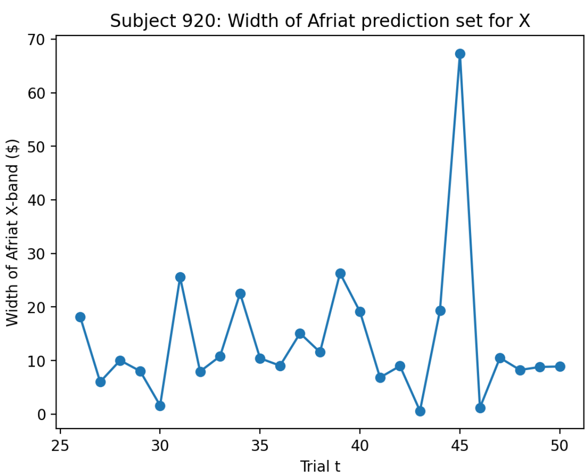
Figure 8: L^1 error vs counterfactual price dispersion: baseline vs narrative.



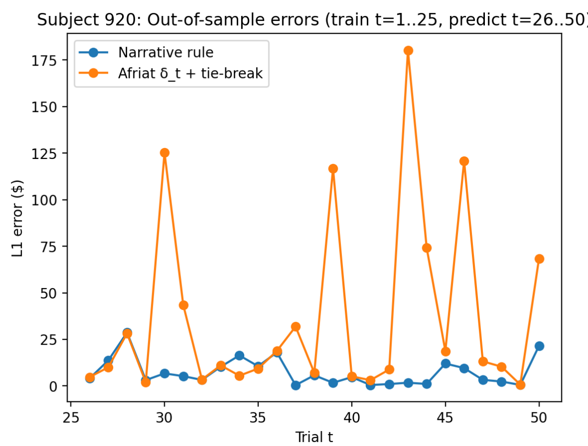
(a) Afriat feasible set for X



(b) X band vs truth (t=50)



(c) X band widths



(d) Per-period L^1 errors

Figure 9: Subject 920: Afriat set vs realized behavior (PapayaTech / X).