Control costs, rational inattention and retail price dynamics

James Costain and Anton Nakov

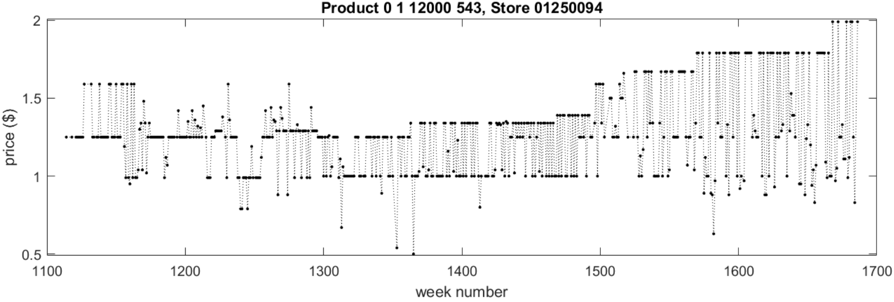
Banco de España, ECB and CEPR

July 2021

Work in progress. Thanks to Luca Riva and Federico Rodari for assistance with the data.

Views expressed here are personal and do not necessarily coincide with our employer's views.

A paradox: “stickiness” but “jumpiness” of retail prices



- A typical supermarket price trajectory is:
  - **Sticky**: it often revisits the exact same nominal values (“price points”)
  - **Volatile**: it frequently makes big jumps between price points
- What causes stickiness+volatility? Does it matter for business cycles?

Existing theories of retail pricing with sales

- Several theories of **sticky plans** with multiple price points
  - **Multiple menu costs**: larger menu cost to change the plan, a smaller one to jump between price points (Eichenbaum et al, '11; Álvarez/Lippi '19)
  - **Stochastic price discrimination**: heterogeneous price elasticities (Varian '80; Guimaraes/Sheedy '11; Kehoe and Midrigan, '15)
  - **"Rational inattention"**: information processing costs cause randomization across discrete price points within a plan (Matejka '16, Stevens '19)
- **Issue:** we don't find much evidence of changes between **plans**.
  Instead: sticky price points that are updated one at a time.

Goal: model sticky+volatile prices, macro implications?

- Present evidence on **sticky price points** vs. **sticky plans**
- Extend our **"control cost" (CC)** model of sticky prices (JMCB, 2019) to generate multiple sticky price points
- Extend **results linking RI with CC** (Steiner/Stewart/Matejka 2017)
  - Define a computable **limited memory RI** framework which **approximates RI**
  - Show that the model generates sticky price points
- Assess importance of price discrimination vs. costly information
- Later: assess importance of this type of pricing for monetary transmission

Some related literature

- State-dependent prices meet microdata
  - Golosov/Lucas '07, Klenow/Malin '10, Midrigan '11, Costain/Nakov '11/'19, Alvarez/Lippi/Paciello '12, Nakamura/Steinsson '15, Cavallo '18
- Stochastic price discrimination and sales
  - Varian '80, Guimaraes/Sheedy '11
- Multiple menu costs and sales
  - Eichenbaum et al '11, Kehoe/Midrigan '15, Alvarez/Lippi '19
- Discrete logit solutions to rational inattention models
  - Matejka/McKay '15, Steiner/Stewart/Matejka '17
- Rational inattention and sticky prices
  - **Matejka '16** (discrete prices), **Stevens '19** (sticky plans)

Two closely related papers

- Stevens '19: sticky plans
  - In her model all prices on the plan change simultaneously; in the data new price points are introduced one at a time
  - Her model is not standard RI – it has two different costs: a fixed cost to obtain full information and replan, plus the usual flow information cost
- Matejka '16: his choice set is the real line
  - Does not specify whether the model is nominal or real
  - If real, then he obtains real stickiness, not nominal
  - If nominal, then nominal payoffs must be stationary to get nominal stickiness
  - Since payoffs are non-stationary in the data, Matejka's model does not apply
  - Our approach: leaving the nominal price unchanged is one possible action

Outline

1. **Microdata**
   - Hard to find evidence of shifts across plans; instead sticky price points
2. **Theory**
   - CC and RI both give logit solutions
   - Equivalence: RI=CC+optimal default distribution: simplifies solution of RI
   - RI implies discretization of the choice set into price points
3. **Model of sticky price points**
   - Limited memory version of RI is computable (and gives an extra parameter)
   - We define "no change" as a possible action: stationary choice set
4. **Numerical findings**
   - **Stickiness** of price points caused by **costs of information processing**
   - **Volatility** across price points driven mainly by **customer heterogeneity**

1. Microdata on Price Trajectories

A closer look at price trajectories

- At time t, define **backwards and forwards windows** for store i, product j:
  - $\mathcal{B}_T^{i,j,t-1} \equiv$ set of all prices observed from $t - T$ to $t - 1$
  - $\mathcal{F}_T^{i,j,t+1} \equiv$ set of all prices observed from $t + 1$ to $t + T$

- **Classify price changes** $(P_{i,j,t} \neq P_{i,j,t-1})$:
  - **Transitory** if $P_{i,j,t} \notin \mathcal{B}_T^{i,j,t-1}$ and $P_{i,j,t} \notin \mathcal{F}_T^{i,j,t+1}$
  - **Recurrence** if $P_{i,j,t} \in \mathcal{B}_T^{i,j,t-1}$
  - **Introduction** if $P_{i,j,t} \notin \mathcal{B}_T^{i,j,t-1}$ and $P_{i,j,t} \in \mathcal{F}_T^{i,j,t+1}$

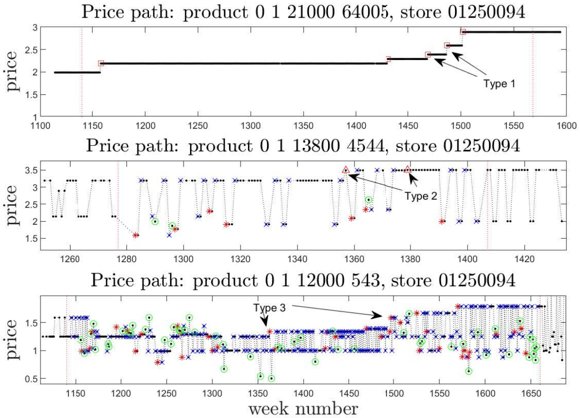
- Further **classify price introductions**:
  - **Type 1:** $\mathcal{B}_T^{i,j,t-1} \cap \mathcal{F}_T^{i,j,t+1} = \emptyset$, and **no recurrences** in $\mathcal{B}_T^{i,j,t-1}$ nor $\mathcal{F}_T^{i,j,t+1}$
    - Typical of *single price policies*
  - **Type 2:** $\mathcal{B}_T^{i,j,t-1} \cap \mathcal{F}_T^{i,j,t+1} = \emptyset$, and **recurrences occur** in $\mathcal{B}_T^{i,j,t-1}$ or $\mathcal{F}_T^{i,j,t+1}$
    - Typical of *sticky plans*
  - **Type 3:** $\mathcal{B}_T^{i,j,t-1} \cap \mathcal{F}_T^{i,j,t+1} \neq \emptyset$.
    - Typical of *sticky price points*

Products with frequent type 1/2/3 introductions: US IRi

Type 1 intros:

Type 2 intros:

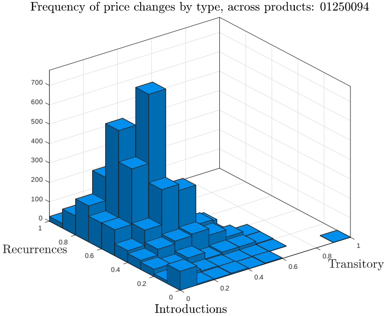
Type 3 intros:

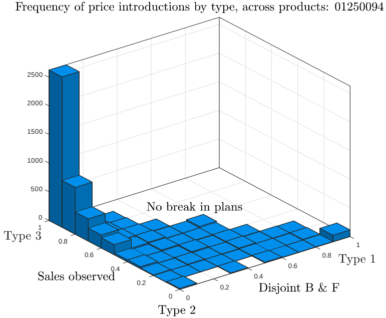


Classifying price changes:		*Classifying introductions:*	
Transitory:	green	Type 1:	squares
Recurrences:	blue	Type 2:	triangles
Introductions:	red	Type 3:	stars

Frequencies of different price change events

Figure: Histograms of event frequencies, across products (example: store 01250094).





Each point on the simplex shows frequency of transitory/recurrences/introductions for a given product. Height of bars represents number of products in each bin.

→ **Many price changes are recurrences.**

Each point on the simplex shows frequency of type 1/2/3 introductions for a given product. Height of bars represents number of products in each bin.

→ **Most introductions are Type 3.**

2. Theory: Costly Decision-Making

Control costs: precise decisions are costly

- **Making a decision** means **allocating probability** $\pi(a)$ over the set of **feasible actions** $a \in \mathcal{A}$.

- **Control cost (CC)** models are **full information** decisions where **precision is costly**:

$$V(\theta) = \max_{\pi \in \Delta(\mathcal{A})} E_\pi u(a, \theta) - \kappa \mathcal{D}(\pi || \eta). \quad (1)$$

- Here we assume **precision** is measured by **relative entropy**:

$$\mathcal{D}(\pi || \eta) \equiv \sum_{a \in \mathcal{A}} \pi(a) \ln \left(\frac{\pi(a)}{\eta(a)} \right).$$

- Calibration requires an **exogenous "benchmark"** action distribution $\eta(a)$ that applies if no decision costs are paid.

Control costs imply weighted logits

- **Expand out** the CC problem:

$$V(\theta) \quad = \quad \max_{\pi(a)} \sum_{a\in\mathcal{A}} \pi(a)u(a,\theta) - \kappa \sum_{a\in\mathcal{A}} \pi(a)\ln\left(\frac{\pi(a)}{\eta(a)}\right)$$
$$\text{s.t.} \quad \sum_{a\in\mathcal{A}} \pi(a) = 1 \quad \text{and} \quad \pi(a) \geq 0, \;\; \forall a \in \mathcal{A}. \tag{2}$$

- FOC for $\pi(a)$:

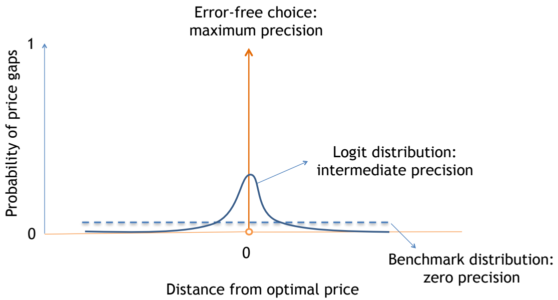
$$u(a,\theta) - \kappa\left(1 + \ln\left(\frac{\pi(a)}{\eta(a)}\right)\right) - \mu \; = \; 0,$$

- Yields a **weighted multinomial logit** (Mattsson/Weibull '02):

$$\pi(a|\theta) \; = \; \frac{\eta(a)\exp(\kappa^{-1}u(a,\theta))}{\sum_{a'\in\mathcal{A}}\eta(a')\exp(\kappa^{-1}u(a',\theta))}\,. \tag{3}$$

  - Here u may be a static payoff, or it may be the **continuation value** of a **dynamic problem**

Nested logits: sequencing of choices is irrelevant



Proposition 1.

Sequence of decisions is irrelevant in dynamic problem:

- *Choose probability of adjustment, then choose distribution over new prices*
- *Choose distribution over new prices, then choose probability to adjust*
- *Choose simultaneously*

Link to Prop. 1 details

Rational inattention: information is costly

- **Rational inattention (RI)** models interpret decision costs in a more specific way, as a **cost of information**:

$$U(\pi_\theta(\theta)) = \max_{\pi(a|\theta)\in\Delta(\mathcal{A})} E_{\pi(a,\theta)} u(a,\theta) - \kappa\mathcal{I}(a;\theta) \tag{4}$$

- Here costs depend on **mutual information** $I(a,\theta)$ between the action a and the state θ:

$$\mathcal{I}(a,\theta) \equiv \mathcal{D}(p(a,\theta)||p_a(a)p_\theta(\theta)) = \sum_{a\in\mathcal{A}}\sum_{\theta\in\Theta}\pi(a,\theta)\ln\left(\frac{\pi(a,\theta)}{\pi_a(a)\pi_\theta(\theta)}\right)$$

  - Mutual information is a **special case** of relative entropy
  - Mutual information measures **deviation from independence**

RI endogenizes the weights in the logit

- Rewrite the **RI problem** in expanded form:

$$U(\pi_\theta(\theta)) = \max_{\pi(a|\theta)\in\Delta(\mathcal{A})} \sum_{\theta\in\Theta} \pi_\theta(\theta) \sum_{a\in\mathcal{A}} \pi(a|\theta) \left[u(a,\theta) - \kappa \ln\left(\frac{\pi(a|\theta)}{\pi_a(a)}\right)\right]$$

$$\text{s.t. } \pi_a(a) = \sum_{\theta\in\Theta} \pi(a|\theta)\pi_\theta(\theta). \tag{5}$$

- Notice:
  - the **RI problem** is just an expectation across various **CC problems!**
  - $\eta(a) = \pi_a(a)$: optimal "benchmark" is the **unconditional action distribution**

- Therefore the solution is:

$$\pi(a|\theta) = \frac{\pi_a(a)\exp(\kappa^{-1}u(a,\theta))}{\sum_{a'\in\mathcal{A}} \pi_a(a')\exp(\kappa^{-1}u(a',\theta))}, \tag{6}$$

$$\pi_a(a) = \sum_{\theta\in\Theta} \pi(a|\theta)\pi_\theta(\theta). \tag{7}$$

Equivalence between RI and CC: static case

Proposition 2.

(Matejka and McKay, 2015). **RI problem (4)-(5)** *represents an expectation across* **CC problems** *under an optimally-chosen benchmark distribution:*

$$U(\pi_\theta(\theta)) = \max_{q(a)\in\Delta(\mathcal{A})} \sum_{\theta\in\Theta} \pi_\theta(\theta) \max_{\pi(a|\theta)\in\Delta(\mathcal{A})} \sum_{a\in\mathcal{A}} \pi(a|\theta) \left[u(a,\theta) - \kappa \ln\left(\frac{\pi(a|\theta)}{q(a)}\right)\right] \tag{8}$$

Therefore, the RI problem is solved by a **weighted multinomial logit**, *with weights equal to the* **marginal probabilities** *of each action:*

$$\pi(a|\theta) = \frac{q(a)\exp(\kappa^{-1}u(a,\theta))}{\sum_{a'\in\mathcal{A}} q(a')\exp(\kappa^{-1}u(a',\theta))}, \tag{9}$$

$$q(a) = \sum_{\theta\in\Theta} \pi(a|\theta)\pi_\theta(\theta). \tag{10}$$

- (9): conditional action probabilities are weighted logits
- (10): optimal weights are unconditional action frequencies

Equivalence between RI and CC: Full and Limited Memory

Proposition 3.

(Steiner, Stewart, and Matejka, 2017). A **dynamic RI problem** *with memory of all previous actions* a^{t-1} *represents an expectation across* **dynamic CC problems** *under an optimally-chosen benchmark distribution* $q(a_t|a^{t-1})$ *at each possible information set* a^{t-1}.

▸ Link to proposition details

Proposition 4.

Short-term memory rational inattention (STMRI)*: A* **dynamic RI problem** *with memory of actions* $\mathcal{B}_\tau^{t-1}$ *in the last* τ *periods only represents an expectation across* **dynamic CC problems** *under an optimally-chosen benchmark distribution* $q(a_t|\mathcal{B}_\tau^{t-1})$ *at each limited information set* $\mathcal{B}_\tau^{t-1}$.

▸ Link to proposition details

Inferring logit weights from data

- Weights on action $a \in \mathcal{A}$ represent average frequencies conditional on the information set a^{t-1}:

$$q(a|a^{t-1}) = \sum_{\theta^t} \pi(a|\theta^t, a^{t-1})\pi(\theta^t|a^{t-1})$$

  - **Not feasible** to infer $q(a|a^{t-1})$ from the data, because we can't repeatedly observe each possible information set a^{t-1}.
  - Exact RI model is hard to identify and hard to compute

- But it is simple to **infer the weights on average**:

$$\bar{q}(a) = \sum_{\theta^t} \sum_{a^{t-1}} \pi(a|\theta^t, a^{t-1})\pi(\theta^t, a^{t-1}) \approx \text{sample frequency of } a$$

  - **Approximate RI model based on average weights** is easy to identify and is feasible to compute

To infer weights, define a stationary action space

- Modelling the action space in a stationary way is crucial
- **Example**: Event $P_{i,j,t} =$ €3.69 may be rare in the dataset, and may vary a lot across product, store, time period, and history of actions.
  - Hard to infer $\pi(P_{i,j,t} =$ €3.69) directly from the data.
- **Example**: Event $P_{i,j,t} \neq P_{i,j,t-1}$ is common in the dataset, and may not vary much with product, store, time period, and history of actions.
  - Might infer $\pi(P_{i,j,t} \neq P_{i,j,t-1})$ directly from the data.

Discreteness of RI solution

- Necessary conditions for RI action distribution:

$$\pi(a|\theta) = \frac{q(a)\exp(\kappa^{-1}u(a,\theta))}{\sum_{a'\in\mathcal{A}} q(a')\exp(\kappa^{-1}u(a',\theta))},$$
$$q(a) = \sum_{\theta\in\Theta} \pi_\theta(\theta)\pi(a|\theta).$$

- But **not all actions** a are chosen with positive probability.
  - Some actions a may have $q(a) = \pi(a|\theta) = 0$ for all θ.

Proposition 5.

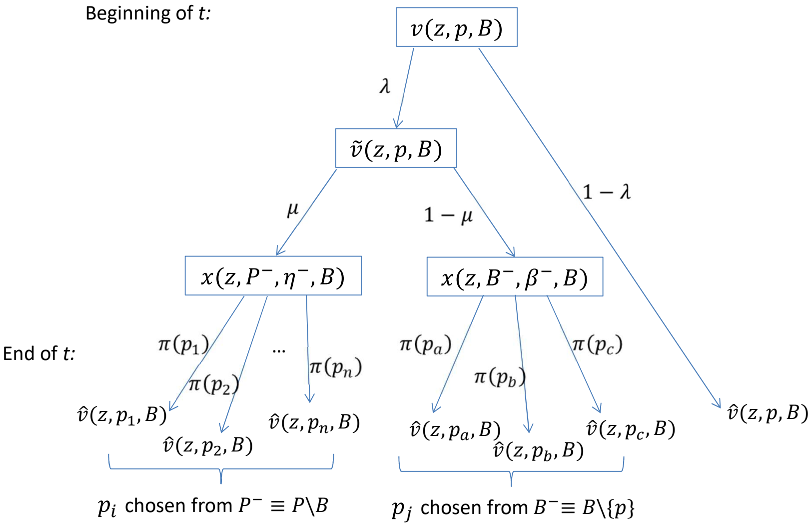
(Fix, 1978). The benchmark distribution $q(a) \geq 0$ solves (5) if it satisfies

$$F(a) \equiv \frac{\sum_{\theta\in\Theta}\pi_\theta(\theta)\sum_{a\in\mathcal{A}}\exp(\kappa^{-1}u(a,\theta))}{\sum_{a'\in\mathcal{A}} q(a')\exp(\kappa^{-1}u(a',\theta))} \leq 1, \quad (11)$$

at all $a \in \mathcal{A}$, with equality wherever $q(a) > 0$ strictly.

- Must check (11) to see which actions have nonzero probability.

CC model with memory: multiple sticky price points



CC model with memory: multiple sticky price points

- Let $\mathcal{P}$ be a large finite set of *real* prices, and let η be a distribution over $\mathcal{P}$.
- Let $\mathcal{B}_t^T$ be the *real* values of the T **most recent nominal prices**

- Value v of **option to adjust**:

$$v(p,z,\mathcal{B}^T) = \max_{\lambda\in[0,1]} \lambda\tilde{v}(p,z,\mathcal{B}^T) + (1-\lambda)\hat{v}(p,z,\mathcal{B}^T) - \kappa w \mathcal{D}(\lambda||\bar{\lambda}) \quad (12)$$

- Value $\tilde{v}$ of option to **choose from $\mathcal{B}^T$ or not**:

$$\tilde{v}(p,z,\mathcal{B}^T) = \max_{\mu\in[0,1]} \mu x(z,\mathcal{P}^-,\eta^-) + (1-\mu)x(z,\mathcal{B}^-,\beta^-) - \kappa w \mathcal{D}(\mu||\bar{\mu}) \quad (13)$$

$$\text{where}: \quad \mathcal{P}^- = \mathcal{P}\diagdown\mathcal{B}^T, \text{ and } \quad \eta^-(\tilde{p}) = \frac{\eta(\tilde{p})}{1-\eta(\mathcal{B}^T)}, \quad \forall\tilde{p}\in\mathcal{P}^- \quad (14)$$

$$\mathcal{B}^- = \mathcal{B}^T\diagdown\{p\}, \text{ and } \quad \beta^-(\tilde{p}) = \frac{\#(\tilde{p}\in\mathcal{B}^T)}{T-\#(p\in\mathcal{B}^T)}, \quad \forall\tilde{p}\in\mathcal{B}^- \quad (15)$$

Link to neural nets algorithm

CC model with memory: multiple sticky price points

- Value x of **setting a new price**:

$$x(z, \mathcal{X}, \xi) \ = \max_{\pi \in \Delta(\mathcal{X})} \sum_{\tilde{p} \in \tilde{\mathcal{X}}} \pi(\tilde{p}) \hat{v}(\tilde{p}, z, \mathcal{B}^T) - \kappa w \mathcal{D}(\pi || \xi) \tag{16}$$

- Value $\hat{v}$ of **producing at price** p:

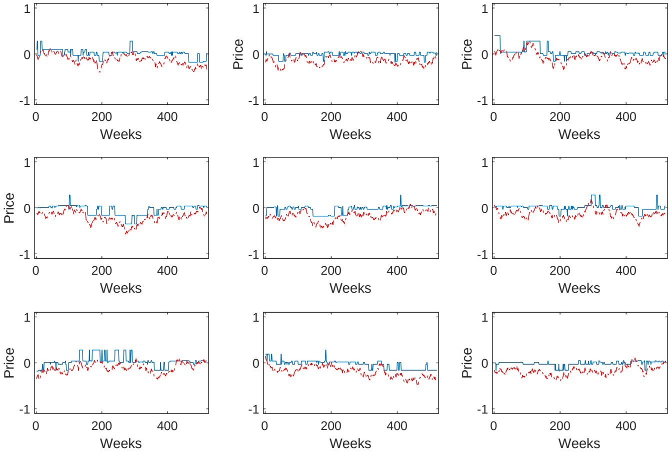
$$\hat{v}(p, z, \mathcal{B}^T) \ = u(z, p) + \frac{1}{1+r} \sum_{z'} \pi_z(z'|z) v(p - i, z', \mathcal{B}^{T\prime}) \tag{17}$$

- **Updating price points** $\mathcal{B}^T \equiv \{b_1, b_2, \ldots b_T\}$ as follows:

$$\mathcal{B}^T \equiv \{p - i,\ b_1 - i,\ b_2 - i, \ldots b_{T-1} - i\}$$

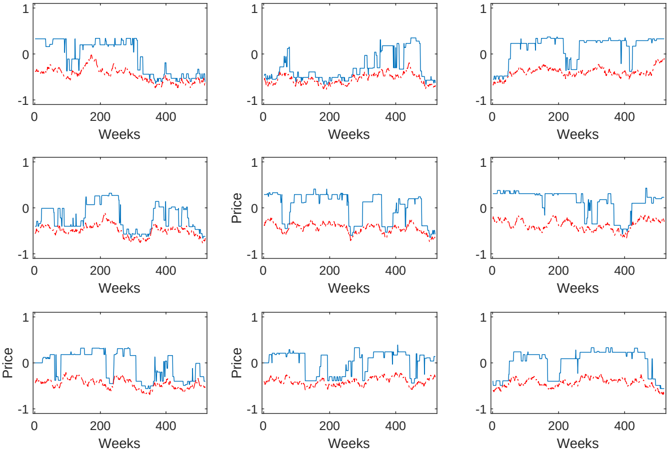
Link to neural nets algorithm

Simulated price trajectories: $\epsilon = 7$ with memory



- Lots of stickiness
- Sometimes previous price points are repeated

Simulated price trajectories: $\epsilon \in \{3, 11\}$ with memory



- Frequent repetition of previous price points
- Markups rise when marginal cost rises

Some summary statistics and conclusions

Table: Model specifications and simulation results (medians)

Parameterizations:	$\epsilon = 7$ No mem.	$\epsilon = 7$ Mem.	$\epsilon = 3, 11$ No mem.	$\epsilon = 3, 11$ Mem.
Price changes				
Adj. frequency (weekly)	0.155	0.138	0.157	0.152
Classifying price changes				
Frequency of recurrences	33.3	67.2	40.0	66.2
Freq. of type 3 introductions	75.0	90.9	75.7	93.6
Short-run volatility ratios				
Ratio VR^{avg}	0.98	0.95	0.85	1.33
Ratio VR^{diff}	16.5	17.5	14.1	23.3
Ratio VR^{reg}	1.52	1.68	1.36	2.36
Ratio VR^{abs}	6.8	7.2	5.6	10.4

- Decision costs with **finite memory** helps explain **stickiness** (frequent recurrences and frequent Type 3 introductions)
- **Heterogeneous demand** (elasticity 3 vs. 11) helps explain excess **volatility**

Some summary statistics and conclusions

Table: Model specifications and simulation results (medians)

Parameterizations:	$\epsilon = 7$ No mem.	$\epsilon = 7$ Mem.	$\epsilon = 3, 11$ No mem.	$\epsilon = 3, 11$ Mem.
Price changes				
Adjustment freq. (weekly)	0.155	0.138	0.157	0.152
Classifying price changes				
Frequency of recurrences	33.3	67.2	40.0	66.2
Freq. of type 3 introductions	75.0	90.9	75.7	93.6
Short-run volatility ratios				
Ratio $VR^{avg}(\tau)$	0.98	0.95	0.85	1.33
Ratio $VR^{diff}(\tau)$	16.5	17.5	14.1	23.3
Ratio $VR^{reg}(\tau)$	1.52	1.68	1.36	2.36
Ratio $VR^{abs}(\tau)$	6.8	7.2	5.6	10.4

- Decision costs with **finite memory** helps explain **stickiness** (frequent recurrences and frequent Type 3 introductions)
- **Heterogeneous demand** (elasticity 3 vs. 11) helps explain excess **volatility**

Some summary statistics and conclusions

Table: Model specifications and simulation results (medians)

Parameterizations:	$\epsilon = 7$ No mem.	$\epsilon = 7$ Mem.	$\epsilon = 3, 11$ No mem.	$\epsilon = 3, 11$ Mem.
Price changes				
Adjustment freq. (weekly)	0.155	0.138	0.157	0.152
Classifying price changes				
Frequency of recurrences	33.3	67.2	40.0	66.2
Freq. of type 3 introductions	75.0	90.9	75.7	93.6
Short-run volatility ratios				
Ratio $VR^{avg}(\tau)$	0.98	0.95	0.85	1.33
Ratio $VR^{diff}(\tau)$	16.5	17.5	14.1	23.3
Ratio $VR^{reg}(\tau)$	1.52	1.68	1.36	2.36
Ratio $VR^{abs}(\tau)$	6.8	7.2	5.6	10.4

- Decision costs with **finite memory** helps explain **stickiness** (frequent recurrences and frequent Type 3 introductions)
- **Heterogeneous demand** (elasticity 3 vs. 11) helps explain excess **volatility**

Conclusions

- **Explaining nominal stickiness:**
  - **Decision costs** help explain intermittent nominal price adjustment
  - **Finite memory** helps explain recurrence of previous nominal prices
  - **Price discrimination** needed to explain excess short-run volatility
- **Computing models of costly decisions**
  - In general, rational inattention models are hard to compute
  - We defined **short-term memory rational inattention (STMRI)**
  - Showed how STMRI can be computed using neural networks
  - Defining a **realistic, stationary action set** makes it possible to identify model parameters
- **RI can only be tested jointly** with a hypothesis about the nature of the action set
  - With an arbitrary action set, RI does not generate nominal stickiness
  - **STMRI, on a realistic action set, can explain nominal stickiness,** including retail sales behavior

Thanks for your (costly) attention!

APPENDIX: Detailed propositions

Invariance to decision sequence

- A convenient fact about the **relative entropy cost function** is that **decisions are invariant to changes in sequencing**, if benchmark distributions are appropriately defined.

Proposition 1.

Let $V(\theta; \mathcal{A}, \eta)$ be the value of (1) with choice set $\mathcal{A}$ and benchmark distribution η.

Split up the action set $\mathcal{A}$ into a **partition** $\mathcal{A}^0 \equiv \{\mathcal{B}_1, \mathcal{B}_2 ..., \mathcal{B}_n\}$.
Define $\eta^0(\mathcal{B}_i) \equiv \sum_{a \in \mathcal{B}_i} \eta(a)$, and $\xi_i(a) \equiv \eta(a)/\eta^0(\mathcal{B}_i)$ for each $a \in \mathcal{B}_i$.

Consider the **two-step CC problem***:*

$$V^0(\theta; \mathcal{A}^0, \eta^0) \quad = \quad \max_{\pi^0 \in \Delta(\mathcal{A}^0)} \; E_{\pi^0} V(\theta; \mathcal{B}, \xi) - \kappa \mathcal{D}(\pi^0 || \eta^0). \tag{18}$$

Two-step problem (18) has the same solution as the one-step problem (1).

▸ Back to main presentation

Equivalence between RI and CC: dynamic case

Proposition 3.

(Steiner, Stewart, and Matejka, 2017). Consider a dynamic RI problem:

$$U(a^0) = \max_{\pi(a_t|\theta^t, a^{t-1}) \in \Delta(\mathcal{A})} E\left[\sum_{t=1}^{\infty} \delta^t \left(u(a_t, \theta_t) - \kappa \mathcal{I}(a_t, \theta^t | a^{t-1}) \right) \middle| a^0 \right]. \tag{19}$$

(i.) *Problem (19) is equivalent to the following double optimization:*

$$U(a^0) = \max_{\pi, q} E\left[\sum_{t=1}^{\infty} \delta^t \left(u(a^t, \theta^t) - \kappa \log\left(\frac{\pi(a_t|\theta^t, a^{t-1})}{q(a_t|a^{t-1})} \right) \right) \middle| a^0 \right]. \tag{20}$$

(ii.) *Problem (20) represents an expectation across* **full-info CC problems** *under an optimal benchmark distribution q:*

$$U(a^{t-1}) \;=\; \delta \max_{q(a|a^{t-1}) \in \Delta(\mathcal{A})} \sum_{\theta^t} \pi(\theta^t | a^{t-1}) V(\theta^t; a^{t-1}, q). \tag{21}$$

(Continues)

Equivalence between RI and CC: dynamic case (continued)

Proposition 3.

(iii.) *In (21), $V(\theta^t; a^{t-1}, q)$ is the value of a recursive, full-info CC problem:*

$$V(\theta^t; a^{t-1}, q) = \max_{\pi(a_t|\theta^t, a^{t-1}) \in \Delta(\mathcal{A})} \sum_{a_t \in \mathcal{A}} \pi(a_t|\theta^t, a^{t-1}) \Bigg[u(a_t, \theta_t) \dots$$

$$-\kappa \ln \left(\frac{\pi(a_t|\theta^t, a^{t-1})}{q(a_t|a^{t-1})} \right) + \delta \sum_{\theta^{t+1}} \pi(\theta^{t+1}|\theta^t) V(\theta^{t+1}; a^t, q) \Bigg]. \tag{22}$$

(iv.) *Hence, (19) and (22) are solved by a* **weighted multinomial logit***:*

$$\pi(a_t|\theta^t, a^{t-1}) \; = \; \frac{q(a_t|a^{t-1}) \exp(\kappa^{-1} \hat{v}(a_t, \theta^t; a^{t-1}, q))}{\sum_{a' \in \mathcal{A}} q(a'|a^{t-1}) \exp(\kappa^{-1} \hat{v}(a', \theta^t; a^{t-1}, q))}, \tag{23}$$

where

$$\hat{v}(a_t, \theta^t; a^{t-1}, q) \; \equiv \; u(a_t, \theta_t) + \delta \sum_{\theta'} \pi(\theta'|\theta^t) V(\theta'; a^t, q). \tag{24}$$

(v.) *The optimal q is the marginal distribution, conditional on signals observed:*

$$q(a_t|a^{t-1}) \; = \; \sum_{\theta^t} \pi(a_t|\theta^t, a^{t-1}) \pi(\theta^t|a^{t-1}). \tag{25}$$

Back to main presentation

Equivalence between STMRI and CC

Proposition 4.

Consider a **short-term memory rational inattention (STMRI)** *problem:*

$$U(\mathcal{B}_\tau^0) = \max_{\pi(a_t|\theta^t, \mathcal{B}_\tau^{t-1}) \in \Delta(\mathcal{A})} E\left[\sum_{t=1}^{\infty} \delta^t \left(u(a_t, \theta_t) - \kappa \mathcal{I}(a_t, \theta^t | \mathcal{B}_\tau^{t-1})\right) \middle| \mathcal{B}_\tau^0\right]. \quad (26)$$

(a.) *The results of Prop. 3 concerning the unlimited memory RI problem (19) extend to the STMRI problem (26). In particular, (26) is solved by a weighted multinomial logit:*

$$\pi(a_t|\theta^t, \mathcal{B}_\tau^{t-1}) = \frac{q(a_t|\mathcal{B}_\tau^{t-1}) \exp(\kappa^{-1} \hat{v}(a_t, \theta^t; \mathcal{B}_\tau^{t-1}))}{\sum_{a' \in \mathcal{A}} q(a'|\mathcal{B}_\tau^{t-1}) \exp(\kappa^{-1} \hat{v}(a', \theta^t; \mathcal{B}_\tau^{t-1}))}, \quad (27)$$

where

$$q(a_t|\mathcal{B}_\tau^{t-1}) = \sum_{\theta^t} \pi(a_t|\theta^t, \mathcal{B}_\tau^{t-1}) \pi(\theta^t|\mathcal{B}_\tau^{t-1}), \quad (28)$$

and $\hat{v}$ is derived from the value function of a dynamic CC problem, as in (24).

(b.) *As memory increases ($\tau \to \infty$), the probabilities and value functions that solve the STMRI problem (26) converge to the solution of the unlimited memory RI problem (19).*

Back to main presentation

How to calibrate the benchmark parameters?

1. **Exogenous uniform benchmark** (Costain/Nakov '19):
   - Benchmark hazard $\bar{\lambda}$ is a constant to be estimated
   - $\mathcal{P}$ is a uniform grid of log real prices
   - The benchmark distribution η is uniform on $\mathcal{P}$

2. **Empirical RI-CC hybrid:**
   In this paper, we instead set the **benchmarks equal to their sample averages** in the data.
   Consider a subset of products I that appear to follow *single price policies*, that is, they only display *Type-1* price introductions.
   - $\bar{\lambda}_t = \bar{\lambda} =$ sample average adjustment frequency in set I.
   - $\tilde{\eta}_t(p) = \tilde{\eta}(p) =$ average histogram, across products in set I, of newly-set log prices, as deviations from product-specific mean.

How to calibrate the benchmark parameters?

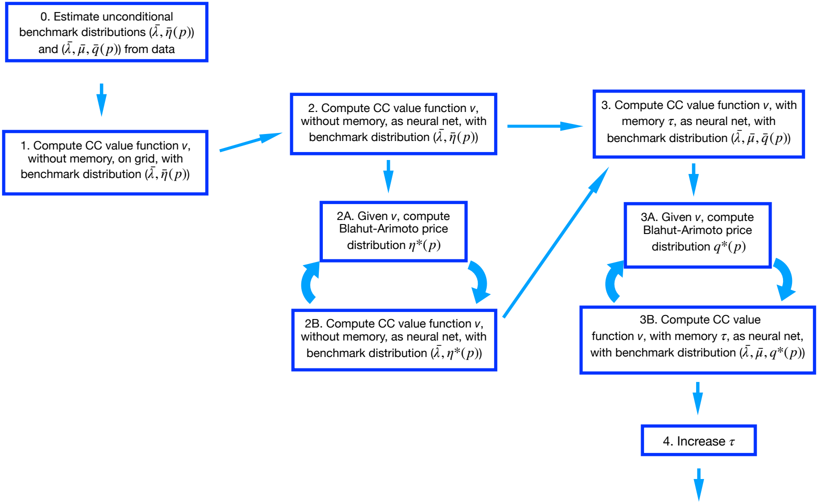
1. **Exogenous uniform benchmark** (Costain/Nakov '19):
   - Benchmark hazard $\bar{\lambda}$ is a constant to be estimated
   - $\mathcal{P}$ is a uniform grid of log real prices
   - The benchmark distribution η is uniform on $\mathcal{P}$

2. **Empirical RI-CC hybrid:**
   In this paper, we instead set the **benchmarks equal to their sample averages** in the data.
   Consider a subset of products I that appear to follow *single price policies*, that is, they only display *Type-1* price introductions.
   - $\bar{\lambda}_t = \bar{\lambda} =$ sample average adjustment frequency in set I.
   - $\tilde{\eta}_t(p) = \tilde{\eta}(p) =$ average histogram, across products in set I, of newly-set log prices, as deviations from product-specific mean.

- To calculate $\tilde{\eta}(p)$:
  1. Find the vector of all new nominal prices $P_{i,j,t}$ chosen for a given product i ;
  2. Calculate the product-specific mean $\bar{P}_i$;
  3. Demean the prices to obtain $p_{i,j,t} = \log(P_{i,j,t}/P_t)$ at all times t such that the price of product i changed at store j;
  4. Aggregate the histograms of prices $p_{i,j,t}$ across all products $i \in I$.

Back to main presentation

Neural nets algorithm: various options



Back to main presentation