Dynamically Aggregating Diverse Information

Annie Liang[1] Xiaosheng Mu[2] Vasilis Syrgkanis[3]

[1]University of Pennsylvania

[2]Princeton [3]Microsoft Research

Introduction

how should an agent acquire information **over time** given

- limited resources, and
- access to multiple kinds of information?

Introduction

how should an agent acquire information **over time** given

- limited resources, and
- access to multiple kinds of information?

examples:

- mayor wants to learn the number of COVID cases in city, allocates limited number of tests across neighborhoods

Introduction

how should an agent acquire information **over time** given

- limited resources, and
- access to multiple kinds of information?

examples:

- mayor wants to learn the number of COVID cases in city, allocates limited number of tests across neighborhoods
- analyst wants to forecast GDP growth, needs to aggregate economic activity across industries and locations

This Talk

- simple model of the dynamic information acquisition problem
- main result: optimal information acquisition strategy can be exactly characterized and has an easily describable structure
- tractability of the model lends itself to application
- show that characterization can be used to derive new results in two settings motivated by particular economic questions:
  - binary choice, competing information providers

Model

Underlying Unknowns

unknown attributes $(\theta_1, \ldots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$

- e.g. each "attribute" is the number of COVID cases in a specific neighborhood
- attributes may be correlated
- learn about θ_i by observing diffusion process X_i^t (more soon)

Underlying Unknowns

unknown attributes $(\theta_1, \ldots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$

- e.g. each "attribute" is the number of COVID cases in a specific neighborhood
- attributes may be correlated
- learn about θ_i by observing diffusion process X_i^t (more soon)

payoff-relevant state: $\omega = \sum_{k=1}^{K} \alpha_k \theta_k$

- e.g. total number of COVID cases in city
- assume weights α_k are known

Attention Allocation

at each $t \in \mathbb{R}_+$, allocate budget of resources across attributes:

- choose $(\beta_1^t, \ldots, \beta_K^t)$ subject to $\beta_1^t + \cdots + \beta_K^t = 1$
- diffusion processes evolve as

$$dX_i^t = \beta_i^t \cdot \theta_i \cdot dt + \sqrt{\beta_i^t} \cdot dB_i^t$$

  where B_i are independent standard Brownian motions.
- more resources $\Rightarrow$ more precise information

Attention Allocation

at each $t \in \mathbb{R}_+$, allocate budget of resources across attributes:

- choose $(\beta_1^t, \ldots, \beta_K^t)$ subject to $\beta_1^t + \cdots + \beta_K^t = 1$
- diffusion processes evolve as

$$dX_i^t = \beta_i^t \cdot \theta_i \cdot dt + \sqrt{\beta_i^t} \cdot dB_i^t$$

  where B_i are independent standard Brownian motions.
- more resources $\Rightarrow$ more precise information

discrete-time analogue: at each time $t \in \mathbb{Z}_+$, choose attention vector $(\beta_1(t), \ldots, \beta_K(t))$ summing to 1, and observe

$$\theta_i + \mathcal{N}\left(0, \frac{1}{\beta_i(t)}\right) \quad \text{for each } i = 1, \ldots, K$$

Decision Problem

- observe complete path of each process
- at each time t the history is $\left\{X_i^{\leq t}\right\}_{i=1}^{K}$
  - **information acquisition strategy** S: map from histories into an attention vector
  - **stopping rule** τ: map from history into decision of whether to stop sampling
- at endogenously chosen end time τ, take action $a \in A$ and receive $u(a, \omega, \tau)$
  - fixing any belief μ about ω at the stopping time, the expected payoff $\max_{a \in A} \mathbb{E}_\mu[u(a, \omega, \tau)]$ is lower for later τ
  - nests geometric discounting, waiting costs $v(a, \omega) - c(\tau)$

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
  - in MAB, actions play the dual role of influencing the evolution of beliefs and determining flow payoffs
  - here they are separated, so information acquisition decisions are driven by learning concerns exclusively

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
  - in MAB, actions play the dual role of influencing the evolution of beliefs and determining flow payoffs
  - here they are separated, so information acquisition decisions are driven by learning concerns exclusively

- recent work on dynamic learning from fixed set of signals:
  - Fudenberg, Strack, and Strzalecki ('18), Che and Mierendorff ('19); Mayskaya ('19); Bardhi ('20); Gossner, Steiner, and Stewart ('20); Azevedo et al ('20)

$\longrightarrow$ we allow many signals with flexible correlation

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
  - in MAB, actions play the dual role of influencing the evolution of beliefs and determining flow payoffs
  - here they are separated, so information acquisition decisions are driven by learning concerns exclusively

- recent work on dynamic learning from fixed set of signals:
  - Fudenberg, Strack, and Strzalecki ('18), Che and Mierendorff ('19); Mayskaya ('19); Bardhi ('20); Gossner, Steiner, and Stewart ('20); Azevedo et al ('20)

  $\longrightarrow$ we allow many signals with flexible correlation

- rational inattention and flexible information acquisition:
  - Steiner, Stewart, and Matejka ('17); Hébert and Woodford ('19); Morris and Strack ('19); Zhong ('19)

  $\longrightarrow$ our signals and information cost are prior-independent

Plan for Talk

1. main result: optimal information acquisition strategy
   - DM initially focuses all attention on one attribute
   - progressively adds in new attributes
   - constant resource/attention allocation during each stage

   focus in talk on $K = 2$ case

2. applications
   - binary choice
   - competing information providers

Main Results

Two Sources

- two unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- agent seeks to learn $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$

Two Sources

- two unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

- agent seeks to learn $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$
- $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

Two Sources

- two unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- agent seeks to learn $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$

- $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

sufficient: $\alpha_1 = \alpha_2$ (so $\omega = \theta_1 + \theta_2$)

Two Sources

- two unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

- agent seeks to learn $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$

- $cov_i := \text{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

sufficient: $\alpha_1 = \alpha_2$ (so $\omega = \theta_1 + \theta_2$), $\Sigma_{12} = \Sigma_{21} \geq 0$ (attributes are positively correlated)

Two Sources

- two unknown attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

- agent seeks to learn $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$

- $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

sufficient: $\alpha_1 = \alpha_2$ (so $\omega = \theta_1 + \theta_2$), $\Sigma_{12} = \Sigma_{21} \geq 0$ (attributes are positively correlated), $\Sigma_{11} = \Sigma_{22}$ (same initial uncertainty)

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

The optimal attention strategy has two stages:

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

The optimal attention strategy has two stages:

1. *At times $t \leq t_1$, DM allocates all attention to attribute 1.*

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$*. Define*

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

The optimal attention strategy has two stages:

1. *At times* $t \leq t_1$*, DM allocates all attention to attribute* 1*.*
2. *At times* $t > t_1$*, DM allocates attention in the constant fraction*

$$(\beta_1^t, \beta_2^t) = \left(\frac{\alpha_1}{\alpha_1 + \alpha_2}, \frac{\alpha_2}{\alpha_1 + \alpha_2}\right).$$

Example 1: Independent Attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

- payoff-relevant state is $\theta_1 + \theta_2$
- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at time $t = 5/6$, posterior covariance matrix is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
  - after, split attention equally

Example 2: Correlated Attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 6 & 2 \\ 2 & 1 \end{pmatrix} \right)$$

- payoff-relevant state is $\theta_1 + \theta_2$
- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at $t = 5/2$, posterior covariance is $\begin{pmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{pmatrix}$
  - after, split attention equally

Interpretation of Strategy

Stage 1

Put all attention on learning about attribute 1, where by assumption $\text{Cov}(\theta_1, \omega) > \text{Cov}(\theta_2, \omega)$.

- start by learning about the attribute that is more informative about the payoff-relevant state
- exclusively acquire information from the source with the higher marginal value

Interpretation of Strategy

- eventually marginal values of the two sources are equal

Interpretation of Strategy

- eventually marginal values of the two sources are equal

Stage 2

Devote attention in constant fraction $\left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)$.

Interpretation of Strategy

- eventually marginal values of the two sources are equal

Stage 2

Devote attention in constant fraction $\left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)$.

- efficient aggregation of information in "prior-free" sense
- these weights produce an unbiased signal about ω:

$$\frac{\alpha_1}{\alpha_1+\alpha_2} \cdot \theta_1 + \frac{\alpha_2}{\alpha_1+\alpha_2} \cdot \theta_2 = \frac{1}{\alpha_1+\alpha_2} \cdot \omega$$

- acquisition of signals in this mixture maintains equivalence of marginal values

Summary of Results for $K > 2$ Attributes

- sufficient condition on prior: $|\Sigma_{ij}| \leq \frac{1}{2K-3} \cdot \Sigma_{ii}, \quad \forall i \neq j.$
  - limits size of covariances relative to variances
  - condition on prior is satisfied at all late times under any reasonable sampling rule
- has nested structure:
  - once DM starts acquiring information from a source, always acquires information from that source
  - progressively adds in new sources
  - at each stage, information acquisition is constant
- information acquisition strategy is again history-independent, depends only on primitives α and Σ

Proof Intuition
(for General K)

Static Problem

- suppose DM takes decision at known time t
- posterior variance of ω at time t can be written $V(q_1, \ldots, q_K)$ where q_i is cumulated attention paid to source i by time t
- any sampling strategy that cumulates to

$$\boldsymbol{q}^*(t) := \underset{q_1,\ldots,q_K \geq 0 \,:\, \sum_k q_k = t}{\operatorname{argmin}} V(q_1, \ldots, q_K)$$

  at time t is optimal (minimizes posterior variance)

Static Problem

- suppose DM takes decision at known time t
- posterior variance of ω at time t can be written $V(q_1, \ldots, q_K)$ where q_i is cumulated attention paid to source i by time t
- any sampling strategy that cumulates to

$$\boldsymbol{q}^*(t) := \underset{q_1,\ldots,q_K \geq 0 \,:\, \sum_k q_k = t}{\operatorname{argmin}} V(q_1, \ldots, q_K)$$

  at time t is optimal (minimizes posterior variance)
- now suppose decision time is uncertain and $q^*(1) = (1/2, 1/2)$ while $q^*(2) = (2, 0)$

  $\longrightarrow$ need to choose between better decision at time $t = 1$ versus $t = 2$ (intertemporal tradeoffs)

Key Idea: Uniformly Optimal Strategies

- suppose these tradeoffs don't exist, i.e. it is possible to achieve $\boldsymbol{q}^*(t)$ at every t
  - such a "uniformly optimal" information acquisition strategy is best for all decision problems

- lemma: uniformly optimal strategy exists iff $\boldsymbol{q}^*(t)$ weakly increases in t in all coordinates.

- in this case, optimal attention levels $\beta^{\boldsymbol{t}}$ are simply the time derivatives of $\boldsymbol{q}^*(t)$

Key Idea: Uniformly Optimal Strategies

- suppose these tradeoffs don't exist, i.e. it is possible to achieve $\boldsymbol{q}^*(t)$ at every t
  - such a "uniformly optimal" information acquisition strategy is best for all decision problems
- lemma: uniformly optimal strategy exists iff $\boldsymbol{q}^*(t)$ weakly increases in t in all coordinates.
- in this case, optimal attention levels β^t are simply the time derivatives of $\boldsymbol{q}^*(t)$
- monotonicity can fail if complementarity/substitution effects across signals are too strong
  - prefer to "take away" attention from one source to improve marginal value of another
  - condition on prior controls size of cross-partials of the posterior variance function V

Step Structure

- at each stage k, agent optimally divides attention among k attributes
- specific mixture of information maintains equivalence of marginal values of those k attributes
- reduces the marginal value of each of the k attributes
- eventually, some new attribute will have the same marginal value as the first k attributes
- at this point the agent expands his observation set to include that new attribute
- repeat reasoning above

Application 1:

Binary Choice

Binary Choice

Drift-Diffusion Model (Ratcliff, 1978)

- two goods with unknown payoffs θ_1 and $-\theta_2$
- agent learns about payoffs before choosing between them
- prior supported on two values $\theta_L < \theta_H$, uncertainty is only over which good is better

Binary Choice

Drift-Diffusion Model (Ratcliff, 1978)

- two goods with unknown payoffs θ_1 and $-\theta_2$
- agent learns about payoffs before choosing between them
- prior supported on two values $\theta_L < \theta_H$, uncertainty is only over which good is better

Fudenberg, Strack, and Strzalecki (2018):

- "uncertain difference" DDM with $(\theta_1, -\theta_2) \sim \mathcal{N}(\mu, \Sigma)$
- both models focus on exogenous learning, but FSS also consider endogenous attention allocation.
- **FSS result:** assume $\Sigma = I$, then optimal attention constant at $(1/2, 1/2)$

Binary Choice

FSS model nested in our setting as case of $\alpha_1 = \alpha_2 = 1$ (given which our characterization holds for all priors)

so can directly apply our $K = 2$ result to generalize FSS's endogenous learning result to all normal priors

$$\begin{pmatrix} \theta_1 \\ -\theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$$

Binary Choice

Corollary

Starting from any prior with $\sigma_1 \geq \sigma_2$, the DM first attends to attribute 1 exclusively, then switches to equal attention at time

$$t_1 = \frac{1/\sigma_2^2 - 1/\sigma_1^2}{1 - \rho^2}.$$

Binary Choice

Corollary

Starting from any prior with $\sigma_1 \geq \sigma_2$*, the DM first attends to attribute* 1 *exclusively, then switches to equal attention at time*

$$t_1 = \frac{1/\sigma_2^2 - 1/\sigma_1^2}{1-\rho^2}.$$

Corollary

Suppose $\sigma_1 \geq \sigma_2$*. Then, holding all else equal:*

- *larger* $\sigma_1 \Rightarrow$ *uniformly more attention towards source* 1
- *larger* $\sigma_2 \Rightarrow$ *uniformly lower attention towards source* 1*;*
- *larger* $|\rho| \Rightarrow$ *uniformly higher attention towards source* 1*.*

Binary Choice

Corollary

Starting from any prior with $\sigma_1 \geq \sigma_2$, the DM first attends to attribute 1 exclusively, then switches to equal attention at time

$$t_1 = \frac{1/\sigma_2^2 - 1/\sigma_1^2}{1 - \rho^2}.$$

Corollary

Suppose $\sigma_1 \geq \sigma_2$. Then, holding all else equal:

- *larger $\sigma_1 \Rightarrow$ uniformly more attention towards source 1*
- *larger $\sigma_2 \Rightarrow$ uniformly lower attention towards source 1;*
- *larger $|\rho| \Rightarrow$ uniformly higher attention towards source 1.*

- see paper for comparative statics in source informativeness

Application 2:
Competing Information Providers

The Game

- each of news sources $i = 1, 2$ has expertise on an unknown θ_i

  (e.g. Biden's popularity with two demographics)

- representative agent wants to learn $\theta_1 + \theta_2$

- imperfect correlation in what the news sources report on:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$$

  (firms are neither monopolists nor perfect competitors)

The Game

- each source chooses the informativeness of news articles
- freely chooses ζ_i (no cost), where one unit of time spent on that source is equivalent to observation of $\theta_i + \mathcal{N}(0, \zeta_i^2)$
- sources want to maximize time spent on their site, payoff is discounted average attention $\int_0^\infty re^{-rt}\beta_i^t \, dt$
- related literature on equilibrium news provision
  - Gentzkow & Shapiro ('08); Chan & Suen ('08, '20), Perego & Yuksel ('18), Galperti & Trevino ('20)

$\longrightarrow$ we focus on role of dynamic considerations

Equilibrium

Proposition

The unique equilibrium is a pure strategy equilibrium (ζ_1^*, ζ_2^*).

- *If* $\sigma_1 = \sigma_2 = \sigma$, *then* $\zeta_1^* = \zeta_2^* = \sigma \cdot \sqrt{\frac{1-\rho}{2r}}$
- *see paper for closed-form expressions in general case*

Equilibrium

Proposition

The unique equilibrium is a pure strategy equilibrium (ζ_1^*, ζ_2^*)*.*

- *If* $\sigma_1 = \sigma_2 = \sigma$*, then* $\zeta_1^* = \zeta_2^* = \sigma \cdot \sqrt{\frac{1-\rho}{2r}}$
- *see paper for closed-form expressions in general case*

On the equilibrium path,

$$(\beta_1(t), \beta_2(t)) = \left(\frac{\zeta_1^*}{\zeta_1^* + \zeta_2^*}, \frac{\zeta_2^*}{\zeta_1^* + \zeta_2^*} \right) \quad \forall t \in \mathbb{R}_+$$

Observations

- there is no “Stage 1” of information gathering: the reader immediately begins mixing in a constant proportion
- despite the possibility of initial asymmetry in how well each attribute is understood
- equilibrium noise levels that exactly offset this prior asymmetry, equalizing marginal values from the beginning

Observations

- there is no “Stage 1” of information gathering: the reader immediately begins mixing in a constant proportion
- despite the possibility of initial asymmetry in how well each attribute is understood
- equilibrium noise levels that exactly offset this prior asymmetry, equalizing marginal values from the beginning
- asymmetry in σ_i does, however, impact how the reader mixes over the sources and the profits that the sources receive.

Firm Profits

Corollary

Equilibrium attention paid to source 1, $\frac{\zeta_1^*}{\zeta_1^* + \zeta_2^*}$,

(a) *exceeds equilibrium attention paid to source 2 if and only if* $\sigma_1 \geq \sigma_2$

(b) *is increasing in* σ_1 *and decreasing in* σ_2

Informativeness of News

Corollary

Equilibrium aggregate noise level, $\zeta_1^* + \zeta_2^*$*,*

(a) *is decreasing in the discount rate* r *(i.e. patience leads to lower quality news)*

(b) *is increasing in* c *if the prior covariance matrix is parametrized as* $c \cdot \Sigma$*;*

(c) *is decreasing in the prior correlation* ρ*.*

Intuition for Role of Patience

- from our $K = 2$ result, we know that for general (ζ_1, ζ_2), attention time path looks like:
  1. exclusive attention to some source i until a time t_i^*
  2. after t_i^*, constant attention $\left(\frac{\zeta_1}{\zeta_1 + \zeta_2}, \frac{\zeta_2}{\zeta_1 + \zeta_2}\right)$
- firms face a tradeoff between
  - optimizing for long-run viewership (encourages higher noise ζ_i)
  - competing to be chosen in the short run (encourages lower ζ_i)
- higher patience $\Longrightarrow$ long run is more important

Conclusion

- we study the problem of dynamic allocation of attention across information sources
- conceptual takeaway:
  - under condition on prior, solution has nested structure, is history-independent, and is robust across decision problems
- practical takeaway:
  - solution very tractable, can use in various applications, e.g. to simplify study of larger model

Possible Extensions

Results hold also for:

- discrete model where agents allocate a fixed budget of precisions each period
- discrete model where agents choose a budget size of precisions each period (at some cost) and allocate it
- intertemporal decision problems (choose actions over time as well, receive payoff that depends on the sequence of actions)

Thank You!