A Random Attention Model

Matias D. Cattaneo  
U. Michigan

Xinwei Ma  
U. Michigan

Yusufcan Masatlioglu  
U. Maryland

Elchin Suleymanov  
U. Michigan
Limited Attention

▷ So many options

▷ People do not pay attention to *all* products
Consideration Set


Consideration Set

The consideration set is made up of products that are taken seriously by the consumer in his or her purchase decision.
Limited Attention

▷ Two-stage Choice

- Masatlioglu, Nakajima, and Ozbay (2012), and many...
Random Attention
Random Attention

MODEL:
- Consumer is able to rank options,
- Consumer chooses the best option in his consideration set,
- Randomness in choices comes from random attention,
Domain and Data

- **Domain**
  - $X = \{a_1, a_2, \ldots, a_K\}$: finite and abstract,

- **Data**
  - $\pi(a_k|S)$: the probability that $a_k$ is chosen from $S \subset X$,

\[
\pi(a_k|S) \geq 0 \quad \text{for all } a_k \in S
\]

\[
\sum_{a_k \in S} \pi(a_k|S) = 1
\]
Random Attention

\[ \mu(T|S): \text{frequency of considering } T \text{ when the feasible set is } S \]

\[ \mu(T|S) \geq 0 \text{ for all } T \subset S \]

\[ \sum_{T \subset S} \mu(T|S) = 1 \]
### Full Attention

<table>
<thead>
<tr>
<th>$S$</th>
<th>${a_1, a_2, a_3}$</th>
<th>${a_1, a_2}$</th>
<th>${a_1, a_3}$</th>
<th>${a_2, a_3}$</th>
<th>${a_1}$</th>
<th>${a_2}$</th>
<th>${a_3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${a_1, a_2, a_3}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>${a_1, a_2}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>${a_1, a_3}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>${a_2, a_3}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Random Attention: Examples

- **Deterministic Attention**

<table>
<thead>
<tr>
<th>$S$</th>
<th>${a_1, a_2, a_3}$</th>
<th>${a_1, a_2}$</th>
<th>${a_1, a_3}$</th>
<th>${a_2, a_3}$</th>
<th>${a_1}$</th>
<th>${a_2}$</th>
<th>${a_3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${a_1, a_2, a_3}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>${a_1, a_2}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>${a_1, a_3}$</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>${a_2, a_3}$</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Random Attention: Examples

Non-Deterministic Attention

\[
S = \{a_1, a_2, a_3\}
\]

\[
T : Consideration \ Sets
\]

<table>
<thead>
<tr>
<th>{a_1, a_2, a_3}</th>
<th>{a_1, a_2}</th>
<th>{a_1, a_3}</th>
<th>{a_2, a_3}</th>
<th>{a_1}</th>
<th>{a_2}</th>
<th>{a_3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>{a_1, a_2, a_3}</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>{a_1, a_2}</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>{a_1, a_3}</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>{a_2, a_3}</td>
<td>0.4</td>
<td>0.6</td>
<td>0.3</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stochastic Choice

\[ \pi(a_k | S) = \sum_{T \subseteq S, \ a_k \text{ is } \succ \text{-best in } T} \mu(T | S) \]

- Unobservables
  - \( \succ \) - complete and transitive
  - \( \mu \) - ????????
Attention Filter


- Consideration set is unchanged when an alternative to which consumer does not pay attention becomes unavailable.

\[ a_k \notin \Gamma(S) \Rightarrow \Gamma(S) = \Gamma(S - a_k) \]

- What is the corresponding condition for \( \mu \)?
Attention Filter

\[ a_k \not\in \Gamma(S) \Rightarrow \Gamma(S) = \Gamma(S - a_k) \]

- Assume \( \Gamma(\{a_1, a_2, a_3\}) = \{a_1, a_2\} \)

| \( \mu(T|S) \) | \{a_1, a_2, a_3\} | \{a_1, a_2\} | \{a_1, a_3\} | \{a_2, a_3\} | \{a_1\} | \{a_2\} | \{a_3\} |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| \{a_1, a_2, a_3\} | 0              | 1              | 0              | 0              | 0              | 0              | 0              |
Attention Filter

\[ a_k \notin \Gamma(S) \Rightarrow \Gamma(S) = \Gamma(S - a_k) \]

- Assume \( \Gamma(\{a_1, a_2, a_3\}) = \{a_1, a_2\} \)
- By the property, \( \Gamma(\{a_1, a_2\}) = \{a_1, a_2\} \) (since \( a_3 \notin \Gamma(\{a_1, a_2, a_3\}) \))

| \( \mu(T|S) \)       | \( \{a_1, a_2, a_3\} \) | \( \{a_1, a_2\} \) | \( \{a_1, a_3\} \) | \( \{a_2, a_3\} \) | \( \{a_1\} \) | \( \{a_2\} \) | \( \{a_3\} \) |
|----------------------|--------------------------|---------------------|---------------------|---------------------|-----------------|-----------------|-----------------|
| \( \{a_1, a_2, a_3\} \) | 0                        | 1                   | 0                   | 0                   | 0               | 0               | 0               |
| \( \{a_1, a_2\} \)     |                           | 1                   | 0                   | 0                   | 0               | 0               | 0               |
Monotonic Attention

Monotonic Attention:

$$\mu(T|S) \leq \mu(T|S - a_k)$$
Monotonic Attention:

\[ \mu(T|S) \leq \mu(T|S - a_k) \]

\[ a_k \not\in \Gamma(S) \Rightarrow \Gamma(S) = \Gamma(S - a_k) \]
Monotonic Attention

Monotonic Attention: \( \mu(T|S) \leq \mu(T|S - a_k) \)

For example, “Uniform Attention”

| \( \mu(T|S) \) | \( \{a_1, a_2, a_3\} \) | \( \{a_1, a_2\} \) | \( \{a_1, a_3\} \) | \( \{a_2, a_3\} \) | \( \{a_1\} \) | \( \{a_2\} \) | \( \{a_3\} \) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| \( \{a_1, a_2, a_3\} \) | 1/7   | 1/7   | 1/7   | 1/7   | 1/7   | 1/7   | 1/7   |
| \( \{a_1, a_2\} \) |       | 1/3   |       |       |       | 1/3   |       |
| \( \{a_1, a_3\} \) |       |       | 1/3   |       |       |       | 1/3   |
| \( \{a_2, a_3\} \) |       |       |       | 1/3   |       |       |       |

- Does \( \mu \) satisfy monotonicity?
Monotonic Attention

Monotonic Attention: $\mu(T|S) \leq \mu(T|S - a_k)$

| $\mu(T|S)$       | $\{a_1, a_2, a_3\}$ | $\{a_1, a_2\}$ | $\{a_1, a_3\}$ | $\{a_2, a_3\}$ | $\{a_1\}$ | $\{a_2\}$ | $\{a_3\}$ |
|-----------------|----------------------|-----------------|-----------------|-----------------|-----------|-----------|-----------|
| $\{a_1, a_2, a_3\}$ | 1/7                  | 1/7             | 1/7             | 1/7             | 1/7       | 1/7       | 1/7       |
| $\{a_1, a_2\}$    | 1/3                  |                 | 1/7             | 1/7             | 1/7       | 1/3       | 1/3       |
| $\{a_1, a_3\}$    |                      | 1/3             |                 | 1/7             | 1/7       | 1/3       | 1/3       |
| $\{a_2, a_3\}$    |                      |                 |                 | 1/3             | 1/7       | 1/3       | 1/3       |
**Monotonic Attention**

\[ \mu(T|S) \leq \mu(T|S - a_k) \]

| \( \mu(T|S) \) | \{a_1, a_2, a_3\} | \{a_1, a_2\} | \{a_1, a_3\} | \{a_2, a_3\} | \{a_1\} | \{a_2\} | \{a_3\} |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \{a_1, a_2, a_3\} | 1/7             | 1/7             | 1/7             | 1/7             | 1/7             | 1/7             | 1/7             |
| \{a_1, a_2\}    |                 | 1/3             |                 |                 |                 |                 |                 |
| \{a_1, a_3\}    |                 |                 | 1/3             |                 |                 |                 |                 |
| \{a_2, a_3\}    |                 |                 |                 | 1/3             |                 |                 |                 |

- **YES**, \( \mu \) is monotonic.
- More examples coming...
Random Attention Model (RAM)

\[ \pi(a_k | S) = \sum_{T \subseteq S, \ a_k \text{ is } \succeq\text{-best in } T} \mu(T | S) \]

- \( \succ \) - complete and transitive
- \( \mu \) - monotonic
Random Attention Model (RAM)

RAM accommodates well-documented and seemingly anomalous behaviors.
**Attraction Effect**

- **Probabilistic Attraction Effect**
  - $a_1$ and $a_2$ are equally chosen in a binary comparison,
  - $a_3$ is a decoy for $a_1$,

| $\pi(a|S)$ | \{a_1, a_2, a_3\} | \{a_1, a_2\} | \{a_1, a_3\} | \{a_2, a_3\} |
|---------|----------------|---------|---------|---------|
| $a_1$   | 1              | 1/2     | 1       |         |
| $a_2$   | 0              | 1/2     |         | 1       |
| $a_3$   | 0              | 0       | 0       | 0       |

$\pi(a_1|X) > \pi(a_1|X - a_3)$
Violation of Regularity

- Random Attention Model allows

\[ \pi(a_k|S) > \pi(a_k|S - a_l) \]

- Removing an alternative can decrease the choice probability
Is the model too general?
The random attention model can be falsified.

For example, the following $\pi$ is outside of the model.

| $\pi(a|S)$ | $\{a_1, a_2, a_3\}$ | $\{a_1, a_2\}$ | $\{a_1, a_3\}$ | $\{a_2, a_3\}$ |
|-----------|----------------------|------------------|------------------|------------------|
| $a_1$     | $1/3$                | 1                | 0                |                  |
| $a_2$     | $1/3$                | 0                |                  | 1                |
| $a_3$     | $1/3$                |                  | 1                | 0                |
RUM vs RAM

- Random Utility satisfies Regularity
- RAM is more general
To illustrate the richness of model, we provide some examples which satisfy monotonicity.
Examples

▶ Fixed Independent Consideration (MM, 2014): Consider a decision maker who pays attention to each alternative with a fixed probability $\gamma$. 
Examples

▷ **Logit Attention:** (BR, 2017) Consider a decision maker who assigns a positive weight, $w_D$, for each non-empty subset of $X$.

$$
\mu(T|S) = \frac{w_T}{\sum_{T' \subset S} w_{T'}}
$$

Psychologically $w_T$ measures the strength associated with the subset $T$. 

Revealed Preference

- How can we deduce preferences under random attention?
- However, richness does not help us much
  - More degree of freedom
  - Allowing many possibilities
  - Less revelations
Multiple Representations

\[(\succ_1, \mu_1), (\succ_2, \mu_2), (\succ_3, \mu_3), \ldots, (\succ_n, \mu_n)\]

**Definition**

\(a_k\) is **revealed to be preferred** to \(a_l\) if \(\succ_i\) ranks \(a_k\) above \(a_l\) for all \(i\).
Revealed Preference

- Finding all possible representations
  - hard and cumbersome

- Need to look for a short-cut
Observation: \( \pi(a_k | S) > \pi(a_k | S - a_l) \) implies “\( a_k \) is better than \( a_l \)”
PROOF:

\[ \pi(a_k | S) = \sum_{T \subset S, a_k \text{ is } \Succbest \text{ in } T} \mu(T | S) \]

\[ = \sum_{a_l \in T \subset S, a_k \text{ is } \Succbest \text{ in } T} \mu(T | S) + \sum_{a_l \notin T \subset S, a_k \text{ is } \Succbest \text{ in } T} \mu(T | S) \]

\[ \leq \sum_{a_l \in T \subset S, a_k \text{ is } \Succbest \text{ in } T} \mu(T | S) + \sum_{a_l \notin T \subset S, a_k \text{ is } \Succbest \text{ in } T} \mu(T | S - a_l) \text{ (by monotonicity)} \]

\[ \leq \sum_{a_l \in T \subset S, a_k \text{ is } \Succbest \text{ in } T} \mu(T | S) + \pi(a_k | S - a_l) \]
PROOF continues...

\[
\pi(a_k|S) - \pi(a_k|S - a_l) \leq \sum_{a_l \in T \subset S, a_k \text{ is } \succ \text{-best in } T} \mu(T|S)
\]

If\( \pi(a_k|S) - \pi(a_k|S - a_l) > 0 \) then there exists at least one \( T \) such that

- \( a_l \in T \)
- \( a_k \text{ is } \succ \text{-best in } T \)
- \( \mu(T|S) \neq 0 \)

Hence, \( a_k \) is revealed to be preferred to \( a_l \). DONE
Revealed Preference

Assume $\pi(a_k|S) > \pi(a_k|S-a_l)$ and $\pi(a_l|S') > p(a_l|S'-a_m)$

- Directly, “$a_k$ is better than $a_l$” and “$a_l$ is better than $a_m$”
- Indirectly, “$a_k$ is better than $a_m$”
Revealed Preference

- $a_k \mathcal{P} a_l$ if $\pi(a_k|S) > \pi(a_k|S-a_l)$

- Let $\bar{\mathcal{P}}$ be the transitive closure of $\mathcal{P}$

- $a_k \bar{\mathcal{P}} a_l$ implies “$a_k$ is better than $a_l$”
Revealed Preference

- While $\bar{P}$ informs us about preference, do we miss some revelation?

**Theorem (Revealed Preference)**

Let $\pi$ have a RAM representation. Then $a_k$ is **revealed to be preferred** to $a_l$ if and only if $a_k \bar{P} a_l$.

- $\bar{P}$ provides all the information we need to know.
A stochastic choice $\pi$ has a RAM representation iff $\bar{\mathcal{P}}$ has no cycle.
Related Literature

- Gul, Natenzon, and Pesendorfer (2014),
- Manzini and Mariotti (2014),
- Echenique, Saito, and Tserenjigmid (2014),
- Brady and Rehbeck (2016),
- Yildiz (2016),
Warp-Up

■ An rich framework,
■ Alternative to RUM,
■ Revealed Preference is a powerful tool,
■ It can be applied
  ● both rational and boundedly rational behavior,
  ● both deterministic and stochastic choice.
Inference and Estimation
Data Generating Process

\[ S \xrightarrow{\mu(T|S)} T \xrightarrow{\succ} a \text{ is } \succ\text{-best in } T \]

Choice Problem \hspace{2cm} Consideration Set \hspace{2cm} Choice

\[ \pi(a|S) \]

**Observable:** choice problem and choice (solid line).
**Unobservable:** attention rule, consideration set and preference (dashed line).
Facts

- Choice rule $\pi(\cdot|\cdot)$ is identified (consistently estimable): we observe the choice problems and choices.

- Attention rule $\mu(\cdot|\cdot)$ is not identified: we do not observe the consideration sets.
  
  - Dimension of $\pi(\cdot|S)$: $|S|$.
  
  - Dimension of $\mu(\cdot|S)$: $2^{|S|} - 1 \gg |S|$.

- In realistic cases, preference $\succ$ is partially identified.

**Definition (Identified Set $\Theta_\pi$)**

$\Theta_\pi$ is the collection of preferences such that

$$\pi(x|S) = \sum_{T \subset S} \mathbb{1}(x \text{ is } \succ\text{-best in } T) \cdot \mu(T|S),$$

for some $\mu$ satisfying monotonicity.
Outline

1 Identification

2 Inference

3 Estimation
Identification: Equivalent Form 1

- \( \pi(x \mid S) > \pi(x \mid S - y) \Rightarrow x > y. \)

- Hard to implement in practice.

- \( \pi \) being estimated makes it more difficult.

**Example (Is \( x > y \)?)**

1. Check \( \pi(x \mid S) \leq \pi(x \mid S - y) \) for all \( S \).

2. Check \( \pi(x \mid S) \leq \pi(x \mid S - z) \) and \( \pi(z \mid S) \leq \pi(z \mid S - y) \) for all \( z \neq x, y \) and \( S \).

3. \( \cdots \) (Check for even longer “chains”)

---

Stochastic Choice Monotonic Attention RAM Identification Inference Estimation
**Identification: Equivalent Form 2**

- $\mathbb{U}$: feasible attention rules, defined by $R\mu \leq 0$.

- Given $\succ$, a mapping from attention rules to choice rules:
  
  $$C_{\succ} : \sum_{T \subset S} \mathbb{1}(k \text{ is } \succ\text{-best in } T) \cdot \mu(T|S).$$

- Hence $\succ \in \Theta_\pi$ if and only if $\pi \in C_{\succ}\mathbb{U} = \{C_{\succ}\mu : R\mu \leq 0\}$.

  $$\mathbb{U} \xrightarrow{C_{\succ}} C_{\succ}\mathbb{U} \ni \pi$$

  **Testing Problem**

- A J-test. Alternatively, can show: $C_{\succ}\mathbb{U} = \{\pi : R_{\succ}\pi \leq 0\}$.

  - Unclear how to construct $R_{\succ}$ from $C_{\succ}$ and $\mathbb{U}$. 
Identification: Equivalent Form 3

- We cannot identify \( \mu \) since
  - Consideration sets are not observed.
  - \( \mu \) has much higher dimension than \( \pi \).

That is, the mapping:
\[
C_{\succ} : \sum_{T \subset S} 1(k \text{ is } \succ\text{-best in } T) \cdot \mu(T|S).
\]
is many-to-one.

- Restrict the class of attention rules.
  - Forward engineering: test if \( \pi \in C_{\succ} \cup \).
  - Backward engineering: (1) construct a attention rule (in the restricted class) from \( \pi \) and \( \succ \); and (2) test if the monotonicity assumption holds.
Example (Triangular Attention Rule)

Take a preference $b \succ c \succ e \succ d \succ a$. $\mu$ is triangular with respect to $\succ$ if, for any menu $S$, it puts weights only on lower contour sets:

$$\{S \cap \{b, c, e, d, a\}, \ S \cap \{c, e, d, a\}, \ S \cap \{e, d, a\}, \ S \cap \{d, a\}, \ S \cap \{a\}\}.$$

Construct a unique $\mu$ from $\pi$. For the previous example,

$$\pi(b|S) = \mu(\{b, c, e, d, a\}|S), \quad \pi(c|S) = \mu(\{c, e, d, a\}|S), \quad \text{etc.}$$

Theorem (Identification)

The following are equivalent:

(1) $\succ \in \Theta_\pi$;

(2) The triangular attention rule constructed from $\pi$ and $\succ$ satisfies the monotonicity assumption.
Main Identification Result

- For each preference $\succ$, we can construct a matrix $\mathbf{R}_\succ$, such that $\succ \in \Theta_\pi$ if and only if $\mathbf{R}_\succ \pi \leq 0$.

- $\mathbf{R}_\succ = \mathbf{RC}_\succ^{-1}$:
  - $\mathbf{C}_\succ^{-1}$: maps choice rules back to triangular attention rules.
  - $\mathbf{R}$: restrictions imposed by the monotonicity assumption.

**Example**

Four elements in the grand set $X = \{a, b, c, d\}$.

1. $\mathcal{S} = \{\{a, b, c, d\}, \{b, c, d\}\}$. Then $b \succ d \succ c \succ a$ gives no constraint.
   $b \succ d \succ a \succ c$ gives one constraint.

2. $\mathcal{S} = \{\{a, b, c, d\}, \{b, c, d\}, \{a, b, c\}, \{b, c\}\}$. Then $b \succ d \succ a \succ c$ gives constraints.
How Many Constraints?

- Five elements: \( X = \{a, b, c, d, e\} \).

<table>
<thead>
<tr>
<th>( S )</th>
<th>Dim. of ( \pi )</th>
<th>Restrictions in ( \mathbb{R}_\succ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>{X}</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>All subsets of size 5 and 4</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>All subsets of size 5, 4 and 3</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>All subsets of size 5, 4, 3 and 2</td>
<td>75</td>
<td>105</td>
</tr>
<tr>
<td>( 2^X \backslash {\emptyset} )</td>
<td>80</td>
<td>131</td>
</tr>
</tbody>
</table>
Outline

1. Identification
2. Inference
3. Estimation
Testing $H_0: \succ \in \Theta_\pi$

- Construct $R_{\succ} = MC_{\succ}^{-1}$

- Data: $\{(y_i, Y_i), \ 1 \leq i \leq N\}$
  
  - $i$’s choice = $y_i \in Y_i = \text{choice problem offered to } i$.

- For each $S$ and each $a \in S$, estimate the choice rule by
  
  $$\hat{\pi}(a|S) = \frac{\sum_{i=1}^{N} 1(Y_i = S, y_i = a)}{\sum_{i=1}^{N} 1(Y_i = S)}$$

- Reject $H_0$ if $R_{\succ} \hat{\pi} \leq 0$ is violated.

- Test statistic: maximum of elements in $R_{\succ} \hat{\pi}$ or 0
  
  $$\mathcal{T} = \sqrt{N} \max \left\{ R_{\succ} \hat{\pi}, \ 0 \right\}.$$
Critical Values

- We need to approximate the distribution of $\mathcal{I}$:

$$\mathcal{I} = \max \left\{ \sqrt{N}R_{\succ} \hat{\pi}, \ 0 \right\}$$

$$= \max \left\{ \sqrt{N}R_{\succ} (\hat{\pi} - \pi) + \sqrt{N}R_{\succ} \pi, \ 0 \right\}.$$  

Asy. Normal \hspace{2cm} Unknown, $\leq 0$ under $H_0$

$$\mathcal{I}^* = \max \left\{ \text{Normal}(0, \hat{V}) + \sqrt{N}\hat{u}, \ 0 \right\}.  \hspace{2cm} \text{Simulate} \hspace{2cm} \text{Plug-in Something}$$

- Then reject $H_0 : \succ \in \Theta_\pi$ if $\mathcal{I} > c_\alpha$.

$$c_\alpha = \inf \left\{ t : \mathbb{P}^*[\mathcal{I}^* > t] \leq \alpha \right\}.$$
Least Favorable Method: $\hat{u} = 0$

Motivation: $H_0: \succ \in \Theta_\pi \Rightarrow R_{\succ} \pi \leq 0$, hence replace by the conservative upper bound 0.

$$T = \max \left\{ \sqrt{N} R_{\succ} (\hat{\pi} - \pi) + \sqrt{N} R_{\succ} \pi, 0 \right\}.$$  

Asy. Normal Unknown, $\leq 0$ under $H_0$

$$T^* = \max \left\{ \text{Normal}(0, \hat{V}) + 0, 0 \right\}.$$  

Simulate Least Favorable

$\succ \in \Theta_\pi \Leftrightarrow R_{\succ} \pi \leq 0$: $T^*$ first order stochastically dominates $T$

$$\lim_N P[T > c_\alpha] \leq \lim_N P^*[T^* > c_\alpha] = \alpha \quad \text{(size control)}.$$  

$\succ \notin \Theta_\pi \Leftrightarrow R_{\succ} \pi \not\leq 0$: $T \xrightarrow{P} +\infty$ while $T^*$ remains bounded in probability

$$\lim_N P[T > c_\alpha] = 1 \quad \text{(consistency)}.$$
(Don’t!) Plug-in Method: \( \hat{u} = (R_{> \hat{\pi}})_- \)

- **(Bad) Motivation:** since \( \pi \) is unknown, replace by its estimate \( \hat{\pi} \).

\[
\mathcal{I} = \max \left\{ \sqrt{N} R_{> \hat{\pi}} (\hat{\pi} - \pi) + \sqrt{N} R_{> \hat{\pi}} \pi, 0 \right\}.
\]

- Asy. Normal
- Unknown, \( \leq 0 \) under \( H_0 \)

\[
\mathcal{I}^* = \max \left\{ \text{Normal}(0, \hat{V}) + \sqrt{N} (R_{> \hat{\pi}})_-, 0 \right\}.
\]

- Simulate
- Plug-in consistent estimate

- **Fails if** \( R_{> \pi} = 0 \)
  - \( \mathbb{P}[\sqrt{N} (R_{> \hat{\pi}})_- \leq 0] \to c > 0. \)
  - With positive probability, \( \mathcal{I} \) dominates \( \mathcal{I}^* \), hence quantiles computed from \( \mathcal{I}^* \) is too small for \( \mathcal{I} \).

- **Also fails if** \( R_{> \pi} < 0 \) but close to 0 (relative to the sample size).
  - Lack of uniformity: for each sample size \( N \), it is possible to find some DGP \( R_{> \pi} < 0 \) such that the rejection probability is strictly larger than \( \alpha \).
Plug-in Method with Shrinkage: \( \hat{u} = \frac{1}{\kappa_N} (R > \hat{\pi})_\sim \)

**Motivation:** shrink the naive plug-in method by a factor \( \kappa_N \to \infty \).

\[
\mathcal{I} = \max \left\{ \sqrt{NR_\sim (\hat{\pi} - \pi)} + \sqrt{NR_\sim \pi}, 0 \right\}.
\]

\[
\mathcal{I}^* = \max \left\{ \text{Normal}(0, \hat{V}) + \frac{\sqrt{N}}{\kappa_N} (R > \hat{\pi})_\sim, 0 \right\}.
\]
Comparison of Three Methods

\[ T = \max \left\{ \sqrt{N}R_\succ (\hat{\pi} - \pi), 0 \right\}. \]

\[ T^* = \max \left\{ \text{Normal}(0, \hat{V}) + \sqrt{N}\hat{u}, 0 \right\}. \]

- \( \succ \in \Theta_\pi \Leftrightarrow R_\succ \pi \leq 0 \)
  - Least Favorable \( \hat{u} = 0 \quad \mathbb{P}[T > c_\alpha] \leq \alpha + O \left( \frac{1}{\sqrt{N}} \right) \)
  - Plug-in \( \hat{u} = (R_\succ \hat{\pi})_- \quad \mathbb{P}[T > c_\alpha] \leq \alpha + O \left( \frac{1}{\sqrt{N}} + 1 \right) \)
  - Shrinkage \( \hat{u} = \frac{1}{\kappa N} (R_\succ \hat{\pi})_- \quad \mathbb{P}[T > c_\alpha] \leq \alpha + O \left( \frac{1}{\sqrt{N}} + \frac{1}{\kappa N} \right) \).

- \( \succ \not\in \Theta_\pi \Leftrightarrow R_\succ \pi \not\leq 0 \)
  - \( (R_\succ \hat{\pi})_- \leq \frac{1}{\kappa N} (R_\succ \hat{\pi})_- \leq 0 \)
  - Power (ability to rule out improbable preferences): PI > S > LF.
Other Methods for Critical Values

- Different choices of $\hat{u}$ are proposed in order to
  - Improve power (rule out improbable preferences) relative to the LF method.
  - Maintain size control (false rejection when $\succ \in \Theta_\pi$).

- Review on testing moment inequalities: ? and ?.
Outline

1 Identification

2 Inference

3 Estimation
Confidence Set

- $\mathcal{C}(\alpha)$ is constructed by inverting the test:

$$\mathcal{C}(\alpha) = \left\{ \succ : \mathcal{T}_\succ \leq c_{\alpha,\succ} \right\}.$$ 

- Covers preferences in the identified set $\Theta_\pi$:

$$\lim_{N} \min_{\succ \in \Theta_\pi} \mathbb{P}\left[ \succ \in \mathcal{C}(\alpha) \right] \geq 1 - \alpha.$$
This Paper

- Generalizes Masatlioglu Nakajima Ozbay (2012) to accommodate random attention scenario.

- Provides conditions under which the preference is partially identified from choice data, without observing consideration sets.

- Constructs test statistics facilitating estimation and inference:
  - Reformulates identification as testing moment inequalities.
    
    There is a large literature on testing moment inequalities and inference in partially identified models.
    
    Other test statistics and methods for critical values can be easily adapted.
  - Provides uniformly valid distributional approximations and critical values.
  - Implements in R and Matlab.