Differential Attention to Attributes in Utility-Theoretic Choice Models

Trudy Ann Cameron
University of Oregon

J.R. DeShazo
School of Public Affairs, UCLA

UNIVERSITY OF OREGON

RAYMOND F. MIKESELL LABORATORY
ENVIRONMENTAL AND RESOURCE ECONOMICS

Outline

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"
- Respondents don't devote enough attention to the different features of the choice set?

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"
- Respondents don't devote enough attention to the different features of the choice set?
- Why not? Attention a scarce resource? Minimal consequences from right or wrong choice?

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"
- Respondents don't devote enough attention to the different features of the choice set?
- Why not? Attention a scarce resource? Minimal consequences from right or wrong choice?
- How might "attention" to a choice task manifest itself, empirically?

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"
- Respondents don't devote enough attention to the different features of the choice set?
- Why not? Attention a scarce resource? Minimal consequences from right or wrong choice?
- How might "attention" to a choice task manifest itself, empirically?
- If attention is "incomplete," what optimizing mechansism might explain patterns of inattention?

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"
- Respondents don't devote enough attention to the different features of the choice set?
- Why not? Attention a scarce resource? Minimal consequences from right or wrong choice?
- How might "attention" to a choice task manifest itself, empirically?
- If attention is "incomplete," what optimizing mechansism might explain patterns of inattention?
- Can we model this, so that it is possible to construct, counterfactually, the full-attention outcome?

Motivation

- Standard criticisms about stated preference survey data
- "People don't take these choice tasks seriously enough, so the choice data are unreliable"
- Respondents don't devote enough attention to the different features of the choice set?
- Why not? Attention a scarce resource? Minimal consequences from right or wrong choice?
- How might "attention" to a choice task manifest itself, empirically?
- If attention is "incomplete," what optimizing mechansism might explain patterns of inattention?
- Can we model this, so that it is possible to construct, counterfactually, the full-attention outcome?
- If attention to different aspects of a choice task can be manipulated, can this be done with malign intent?

Acknowledgements

Evolution of paper

- **Initial idea:** "Dissecting the Random Component," Fifth Triennial Invitational Choice Symposium hosted by UC Berkeley at Asilomar in Pacific Grove, CA, in June 2001.

Acknowledgements

Evolution of paper

- **Initial idea:** "Dissecting the Random Component," Fifth Triennial Invitational Choice Symposium hosted by UC Berkeley at Asilomar in Pacific Grove, CA, in June 2001.
- **Theory section:** "Recent Progress on Endogeneity in Choice Modeling," Sixth Triennial Invitational Choice Symposium in Estes Park, CO in June 2004.

Acknowledgements

Evolution of paper

- **Initial idea:** "Dissecting the Random Component," Fifth Triennial Invitational Choice Symposium hosted by UC Berkeley at Asilomar in Pacific Grove, CA, in June 2001.
- **Theory section:** "Recent Progress on Endogeneity in Choice Modeling," Sixth Triennial Invitational Choice Symposium in Estes Park, CO in June 2004.
- **Empirical example:** "Behavioral Frontiers in Choice Models," Seventh Triennial Invitational Choice Symposium at the Wharton School in Philadelphia, PA, in June 2007

Attention in Choice Models

- Typical assumptions: up to a random component, investigator knows all information that individual uses to make choice–individuals fully attend to, and costlessly process, all the information presented to them within a choice set

Attention in Choice Models

- Typical assumptions: up to a random component, investigator knows all information that individual uses to make choice–individuals fully attend to, and costlessly process, all the information presented to them within a choice set
- However, constituent elements of attention, including cognition and time, are scarce resources which rational individuals should allocate optimally (Simon, 1955; March, 1978; Heiner, 1985; de Palma, et al., 1994; Conlisk, 1996; Gabiax and Laibson, 2000)

Attention in Choice Models

- Typical assumptions: up to a random component, investigator knows all information that individual uses to make choice–individuals fully attend to, and costlessly process, all the information presented to them within a choice set
- However, constituent elements of attention, including cognition and time, are scarce resources which rational individuals should allocate optimally (Simon, 1955; March, 1978; Heiner, 1985; de Palma, et al., 1994; Conlisk, 1996; Gabiax and Laibson, 2000)
- Optimal allocation of attention will depend on marginal benefits, marginal costs of further information processing

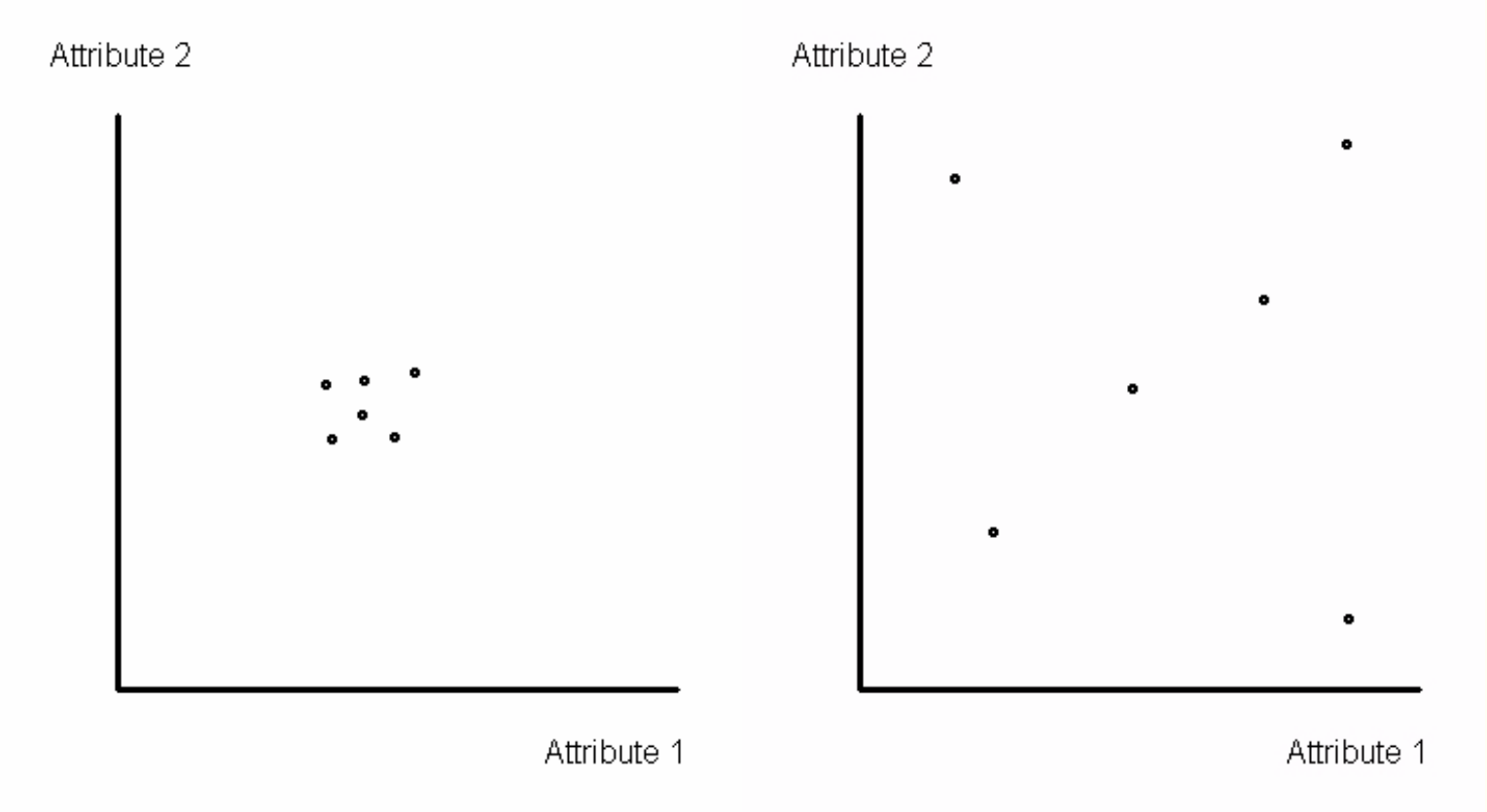
Attention in Choice Models

- Typical assumptions: up to a random component, investigator knows all information that individual uses to make choice–individuals fully attend to, and costlessly process, all the information presented to them within a choice set
- However, constituent elements of attention, including cognition and time, are scarce resources which rational individuals should allocate optimally (Simon, 1955; March, 1978; Heiner, 1985; de Palma, et al., 1994; Conlisk, 1996; Gabiax and Laibson, 2000)
- Optimal allocation of attention will depend on marginal benefits, marginal costs of further information processing
- Thus, prior to making a choice, individual may rationally attend to some attributes/alternatives more than others

Attention in Choice Models

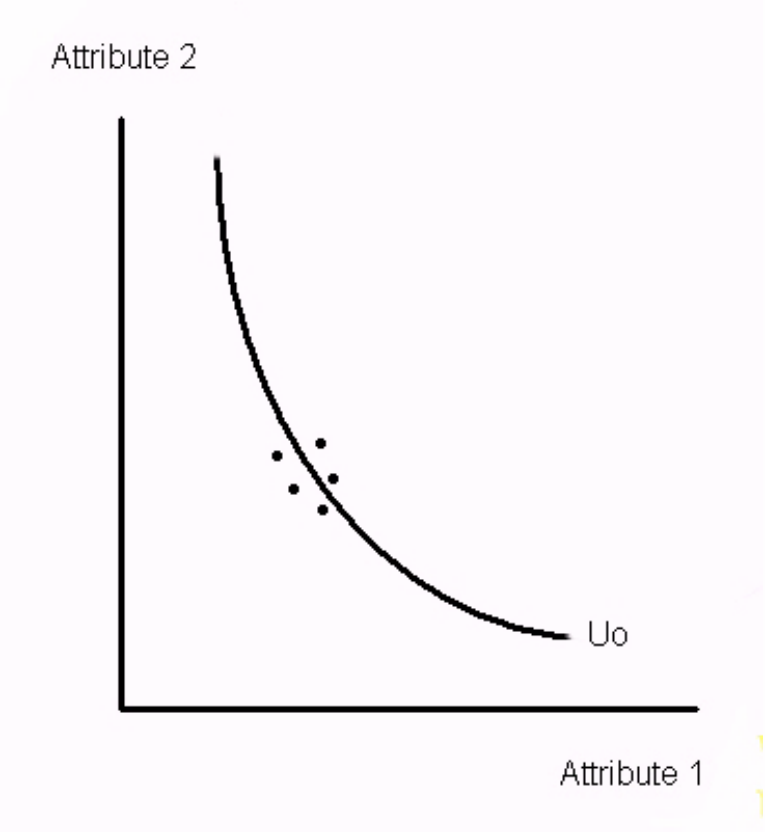
- Typical assumptions: up to a random component, investigator knows all information that individual uses to make choice–individuals fully attend to, and costlessly process, all the information presented to them within a choice set
- However, constituent elements of attention, including cognition and time, are scarce resources which rational individuals should allocate optimally (Simon, 1955; March, 1978; Heiner, 1985; de Palma, et al., 1994; Conlisk, 1996; Gabiax and Laibson, 2000)
- Optimal allocation of attention will depend on marginal benefits, marginal costs of further information processing
- Thus, prior to making a choice, individual may rationally attend to some attributes/alternatives more than others
- This paper – attention to **attributes** (attention to alternatives in a separate paper)

Basics - Similarity in Attribute Space

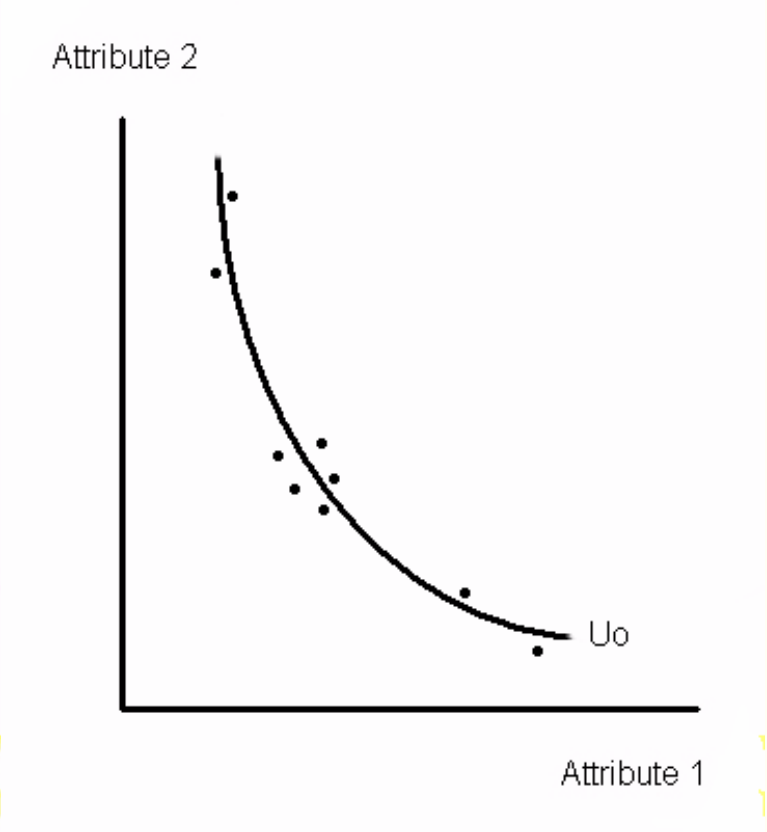


"Similar" in Attribute Space **"Different" in Attribute Space**

Basics - Similarity in Utility Space



"Similar" in attribute space and also in utility level provided



"Different" in attribute space, but similar in utility levels...for someone with *these* preferences

Basics - choice sets

Which attribute of these cars seems most important to consider?




Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

Which attribute of these cars seems most important to consider?




Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

Which attribute of these cars seems most important to consider?




Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

Which attribute of these cars seems most important to consider?




Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

Which attribute of these cars seems most important to consider?




Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

Which attribute of these cars seems most important to consider?




Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

Despite their similarity on non-price attributes, "labels" may be very important




	Honda Accord	Toyota Camry	Chev. Malibu	Pontiac G6
Engine	177-hp 2.4-liter I-4	158-hp 2.4-liter I-4	169-hp 2.4-liter I-4	164-hp 2.4-liter I-4
Transm.	5-spd auto. w/OD	5-spd auto. w/OD	4-spd auto. w/OD	4-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	19 - 21 mpg	17 - 22 mpg	15 - 22 mpg
Fuel Economy Hwy	25 - 31 mpg	28 - 31 mpg	26 - 30 mpg	22 - 30 mpg
MSRP	$20,360	$24,120	$19,345	$29,910

Basics - choice sets

When many attributes differ, you need to carefully consider all of them




	Honda Accord	Honda Civic Hybrid	Honda Odyssey	Honda Ridgeline
Engine	177-hp 2.4-liter I-4	93-hp 1.3-liter I-4 (gasoline hybrid)	241-hp 3.5-liter V-6	247-hp 3.5-liter V-6
Transm.	5-spd auto. w/OD	2-spd CVT w/OD	5-spd auto. w/OD	5-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	40 mpg	16 - 17 mpg	15 mpg
Fuel Economy Hwy	25 - 31 mpg	45 mpg	23 - 25 mpg	20 mpg
MSRP	$20,360	$22,600	$25,860	$29,910

Basics - choice sets

When many attributes differ, you need to carefully consider all of them




	Honda Accord	Honda Civic Hybrid	Honda Odyssey	Honda Ridgeline
Engine	177-hp 2.4-liter I-4	93-hp 1.3-liter I-4 (gasoline hybrid)	241-hp 3.5-liter V-6	247-hp 3.5-liter V-6
Transm.	5-spd auto. w/OD	2-spd CVT w/OD	5-spd auto. w/OD	5-spd auto. w/OD
Fuel Economy City	17 - 22 mpg	40 mpg	16 - 17 mpg	15 mpg
Fuel Economy Hwy	25 - 31 mpg	45 mpg	23 - 25 mpg	20 mpg
MSRP	$20,360	$22,600	$25,860	$29,910

Attention - determinants

Overall attention to a choice problem depends upon three basic factors

- **Ability**: the individual's cognitive budget, defined by cognitive capacity and information-processing ability

Attention - determinants

Overall attention to a choice problem depends upon three basic factors

- **Ability**: the individual's cognitive budget, defined by cognitive capacity and information-processing ability
- **Inclination**: depends upon the importance of the choice to the individual, including its *consequentiality*

Attention - determinants

Overall attention to a choice problem depends upon three basic factors

- **Ability**: the individual's cognitive budget, defined by cognitive capacity and information-processing ability
- **Inclination**: depends upon the importance of the choice to the individual, including its *consequentiality*
  - More attention to choices among houses, new cars, potential spouses

Attention - determinants

Overall attention to a choice problem depends upon three basic factors

- **Ability**: the individual's cognitive budget, defined by cognitive capacity and information-processing ability
- **Inclination**: depends upon the importance of the choice to the individual, including its *consequentiality*
  - More attention to choices among houses, new cars, potential spouses
  - Less attention to choice of hotels, rental cars, or dates

Attention - determinants

Overall attention to a choice problem depends upon three basic factors

- **Ability**: the individual's cognitive budget, defined by cognitive capacity and information-processing ability
- **Inclination**: depends upon the importance of the choice to the individual, including its *consequentiality*
  - More attention to choices among houses, new cars, potential spouses
  - Less attention to choice of hotels, rental cars, or dates
- **Opportunity**: the individual may be time-constrained in making a choice (opportunity cost of time)

Attention - versus marginal utility

We do not normally observe *attention* to a choice problem.

- Time on task? - may be longer if distracted, or longer if inferior cognitive skills
- Subjective attention? - difficult to elicit ("Were you paying attention to that choice?" "Huh? Of course I was!")

Attention - versus marginal utility

We do not normally observe *attention* to a choice problem.

- Time on task? - may be longer if distracted, or longer if inferior cognitive skills
- Subjective attention? - difficult to elicit ("Were you paying attention to that choice?" "Huh? Of course I was!")

We observe only a marginal *effect* of each attribute on choice probabilities, which confounds

- Attention to that attribute
- Marginal utility from that attribute

Attention - versus marginal utility

We do not normally observe *attention* to a choice problem.

- Time on task? - may be longer if distracted, or longer if inferior cognitive skills
- Subjective attention? - difficult to elicit ("Were you paying attention to that choice?" "Huh? Of course I was!")

We observe only a marginal *effect* of each attribute on choice probabilities, which confounds

- Attention to that attribute
- Marginal utility from that attribute

"Don't know...Don't Care?" Zero apparent marginal effect for an attribute can mean

- Respondent *didn't notice* the differences in this attribute across alternatives
- The respondent *did notice* these differences but they have *no effect* on his/her utility

Attention - range

Suppose individual *does* value a particular attribute

Case 1: not resource-constrained; full attention to levels of all attributes

- "true" marginal utilities (MUs) can be estimated

Attention - range

Suppose individual *does* value a particular attribute

Case 1: not resource-constrained; full attention to levels of all attributes

- "true" marginal utilities (MUs) can be estimated

Case 3: heavily resource-constrained; no attention to levels of some attribute

- apparent MU of attribute is zero (hasty choice, didn't think to consider X)

Attention - range

Suppose individual *does* value a particular attribute

Case 1: not resource-constrained; full attention to levels of all attributes

- "true" marginal utilities (MUs) can be estimated

Case 2: somewhat resource-constrained; incomplete attention to levels of some attributes

- apparent MU is attenuated in some or all cases

Case 3: heavily resource-constrained; no attention to levels of some attribute

- apparent MU of attribute is zero (hasty choice, didn't think to consider X)

Attention - consequences

What happens when cognitive resource constraints are binding?

- **Possible Scenario**: individuals pay *proportionately* less attention to every attribute?

Attention - consequences

What happens when cognitive resource constraints are binding?

- **Possible Scenario**: individuals pay *proportionately* less attention to every attribute?
  - *Proportional* attenuation in all marginal utilities
  - Result: No effect on WTP (ratio of marginal utilities)
  - Observationally equivalent to a *scale effect*
  - All MUs proportionately lower $\equiv$ error dispersion larger (i.e. "noisier" choices)

Attention - consequences

What happens when cognitive resource constraints are binding?

- **Possible Scenario**: individuals pay *proportionately* less attention to every attribute?
  - *Proportional* attenuation in all marginal utilities
  - Result: No effect on WTP (ratio of marginal utilities)
  - Observationally equivalent to a *scale effect*
  - All MUs proportionately lower $\equiv$ error dispersion larger (i.e. "noisier" choices)
- **Possible Scenario**: individuals may pay relatively more attention to *price* attribute and relatively less to other attributes?

Attention - consequences

What happens when cognitive resource constraints are binding?

- **Possible Scenario**: individuals pay *proportionately* less attention to every attribute?
  - *Proportional* attenuation in all marginal utilities
  - Result: No effect on WTP (ratio of marginal utilities)
  - Observationally equivalent to a *scale effect*
  - All MUs proportionately lower $\equiv$ error dispersion larger (i.e. "noisier" choices)
- **Possible Scenario**: individuals may pay relatively more attention to *price* attribute and relatively less to other attributes?
  - Less attenuation in the estimated marginal utility of net income (WTP denominator stays large)
  - More attenuation for the marginal utilities of other attributes (WTP numerator shrinks more)
  - Result: Willingness to pay is *biased toward zero*

Related Research

Xavier Gabaix, David Laibson, Guillermo Moloche, AER (2006)

- Mouselab experiment
- Rows of boxes with hidden payouts
- Objective: choose row with highest sum (one trial chosen for payout)

Related Research

Xavier Gabaix, David Laibson, Guillermo Moloche, AER (2006)

- Mouselab experiment
- Rows of boxes with hidden payouts
- Objective: choose row with highest sum (one trial chosen for payout)
- Click on boxes to reveal more terms in sum (Go Fish!)
- Choose which (and how many) boxes to click before choice
- Row amounts have declining variances (fewer surprises)

Related Research

Xavier Gabaix, David Laibson, Guillermo Moloche, AER (2006)

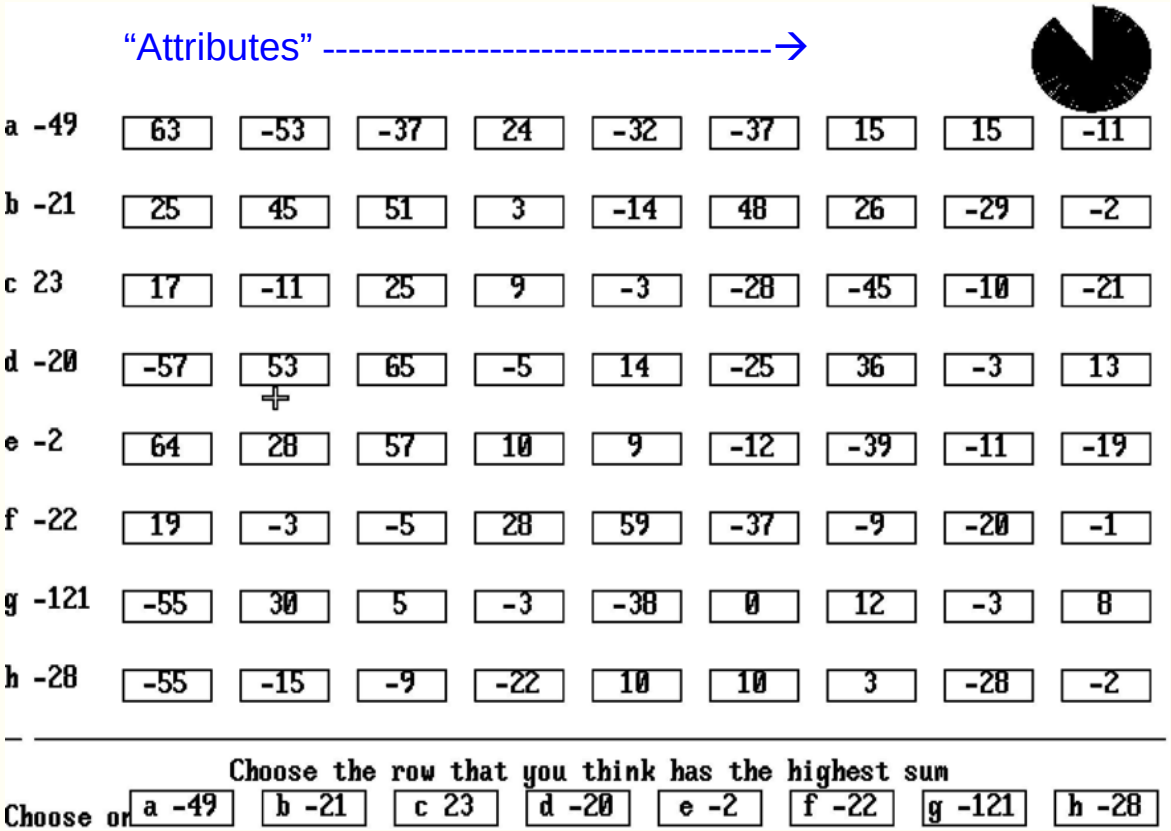
- Mouselab experiment
- Rows of boxes with hidden payouts
- Objective: choose row with highest sum (one trial chosen for payout)
- Click on boxes to reveal more terms in sum (Go Fish!)
- Choose which (and how many) boxes to click before choice
- Row amounts have declining variances (fewer surprises)
- Time-constrained choices
- Directed cognition (i.e. attention), bounded rationality

Related Research

THE AMERICAN ECONOMIC REVIEW SEPTEMBER 2006

Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model

By Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg

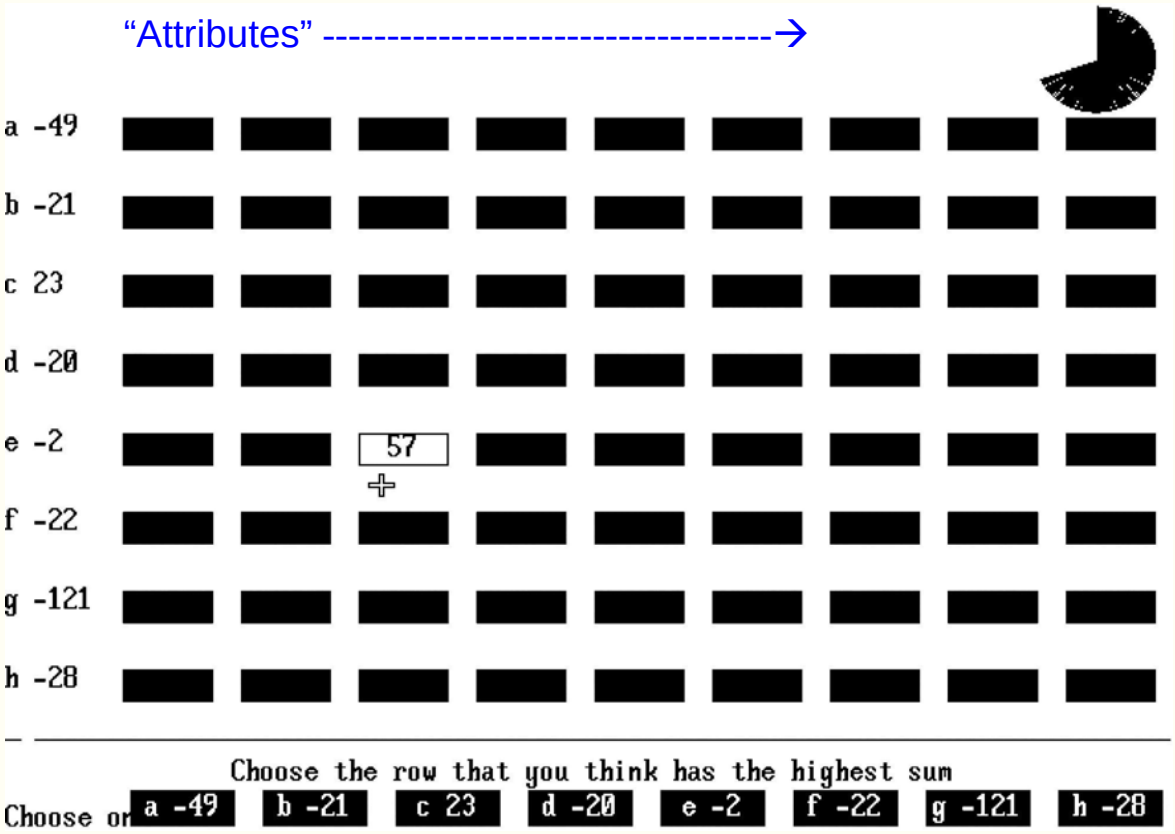


Related Research

THE AMERICAN ECONOMIC REVIEW SEPTEMBER 2006

Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model

By Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg



Related Research

Gabaix, Laibson, Moloche, Weinberg, AER (2006)

- Their insights can be translated to match the two main drivers of attention we identify from our model (in a special case where boxes are revealed one whole column at a time)

Related Research

Gabaix, Laibson, Moloche, Weinberg, AER (2006)

- Their insights can be translated to match the two main drivers of attention we identify from our model (in a special case where boxes are revealed one whole column at a time)
- An *elegant* start on the attention problem, but all of the "attributes" in the experiment are money

Related Research

Gabaix, Laibson, Moloche, Weinberg, AER (2006)

- Their insights can be translated to match the two main drivers of attention we identify from our model (in a special case where boxes are revealed one whole column at a time)
- An *elegant* start on the attention problem, but all of the "attributes" in the experiment are money
- No need to worry about the distinction between attribute-space and utility-space (they are the same)

Related Research

Gabaix, Laibson, Moloche, Weinberg, AER (2006)

- Their insights can be translated to match the two main drivers of attention we identify from our model (in a special case where boxes are revealed one whole column at a time)
- An *elegant* start on the attention problem, but all of the "attributes" in the experiment are money
- No need to worry about the distinction between attribute-space and utility-space (they are the same)
- No need to worry about differing marginal utilities for different attributes

Related Research

Gabaix, Laibson, Moloche, Weinberg, AER (2006)

- Their insights can be translated to match the two main drivers of attention we identify from our model (in a special case where boxes are revealed one whole column at a time)
- An *elegant* start on the attention problem, but all of the "attributes" in the experiment are money
- No need to worry about the distinction between attribute-space and utility-space (they are the same)
- No need to worry about differing marginal utilities for different attributes
- No need to worry about differences across individuals in these attribute marginal utilities

Related Research

Gabaix, Laibson, Moloche, Weinberg, AER (2006)

- Their insights can be translated to match the two main drivers of attention we identify from our model (in a special case where boxes are revealed one whole column at a time)
- An *elegant* start on the attention problem, but all of the "attributes" in the experiment are money
- No need to worry about the distinction between attribute-space and utility-space (they are the same)
- No need to worry about differing marginal utilities for different attributes
- No need to worry about differences across individuals in these attribute marginal utilities
- All of these concerns *are* relevant for real choice problems

Related Research

David Hensher et al. Design-of-designs study (several papers))

- Ask respondents explicitly about which attributes they *ignored* in making their choices (overall, not on a choice-by-choice basis)

Related Research

David Hensher et al. Design-of-designs study (several papers))

- Ask respondents explicitly about which attributes they *ignored* in making their choices (overall, not on a choice-by-choice basis)
- Finding: Number of attributes considered is *lower* when sets of attributes are drawn from distributions with *narrower ranges* (i.e. when alternatives are less different)

Related Research

David Hensher et al. Design-of-designs study (several papers))

- Ask respondents explicitly about which attributes they *ignored* in making their choices (overall, not on a choice-by-choice basis)
- Finding: Number of attributes considered is *lower* when sets of attributes are drawn from distributions with *narrower ranges* (i.e. when alternatives are less different)
- Individual's processing strategies depend on the *nature* of the attribute information, not just the *quantity* of such information

Related Research

David Hensher et al. Design-of-designs study (several papers))

- Ask respondents explicitly about which attributes they *ignored* in making their choices (overall, not on a choice-by-choice basis)
- Finding: Number of attributes considered is *lower* when sets of attributes are drawn from distributions with *narrower ranges* (i.e. when alternatives are less different)
- Individual's processing strategies depend on the *nature* of the attribute information, not just the *quantity* of such information
- Individuals' information processing strategies "should be built into the estimation of choice data from stated choice studies"

Related Research

David Hensher et al. Design-of-designs study (several papers))

- Ask respondents explicitly about which attributes they *ignored* in making their choices (overall, not on a choice-by-choice basis)
- Finding: Number of attributes considered is *lower* when sets of attributes are drawn from distributions with *narrower ranges* (i.e. when alternatives are less different)
- Individual's processing strategies depend on the *nature* of the attribute information, not just the *quantity* of such information
- Individuals' information processing strategies "should be built into the estimation of choice data from stated choice studies"
- *Exactly* what we endeavor to do here

Binary Choice Model

Indirect utility function: linear, additively separable in

- net income (Y_i minus T_i^j, the cost of option j),
- each of several attributes, X_{ki}, k=1,. . . ,K.

Alternative 1: $V_i^1 = \beta_1 \left(Y_i - T_i^1\right) + \sum_{k=2}^{K} \beta_k X_{ki}^1 + \varepsilon_i^1$

Alternative 0: $V_i^0 = \beta_1 \left(Y_i - T_i^0\right) + \sum_{k=2}^{K} \beta_k X_{ki}^0 + \varepsilon_i^0$

Binary Choice Model

Indirect utility function: linear, additively separable in

- net income (Y_i minus T_i^j, the cost of option j),
- each of several attributes, X_{ki}, k=1,. . . ,K.

Alternative 1: $V_i^1 = \beta_1 \left(Y_i - T_i^1\right) + \sum_{k=2}^{K} \beta_k X_{ki}^1 + \varepsilon_i^1$

Alternative 0: $V_i^0 = \beta_1 \left(Y_i - T_i^0\right) + \sum_{k=2}^{K} \beta_k X_{ki}^0 + \varepsilon_i^0$

Utility-difference:

$$V_i^1 - V_i^0 = \beta_1 \left(T_i^0 - T_i^1\right) + \sum_{k=2}^{K} \beta_k \left(X_{ki}^1 - X_{ki}^0\right) + \left(\varepsilon_i^1 - \varepsilon_i^0\right)$$

Binary Choice Model

Indirect utility function: linear, additively separable in

- net income (Y_i minus T_i^j, the cost of option j),
- each of several attributes, X_{ki}, k=1,. . . ,K.

Alternative 1: $V_i^1 = \beta_1 \left(Y_i - T_i^1\right) + \sum_{k=2}^{K} \beta_k X_{ki}^1 + \varepsilon_i^1$

Alternative 0: $V_i^0 = \beta_1 \left(Y_i - T_i^0\right) + \sum_{k=2}^{K} \beta_k X_{ki}^0 + \varepsilon_i^0$

Utility-difference:

$$V_i^1 - V_i^0 = \beta_1 \left(T_i^0 - T_i^1\right) + \sum_{k=2}^{K} \beta_k \left(X_{ki}^1 - X_{ki}^0\right) + \left(\varepsilon_i^1 - \varepsilon_i^0\right)$$

$$= -\beta_1 t_i + \sum_{k=2}^{K} \beta_k x_{ki} + \varepsilon_i$$

where $\left(T_i^0 - T_i^1\right) = \left(X_{1i}^1 - X_{1i}^0\right) = x_{1i} = -t_i$ is treated differently, due to the special role of β_1 in calculating WTP.

Willingness to pay (WTP)

WTP is that program cost which makes the individual just indifferent between

- paying for the program and enjoying its benefits, and
- not paying for the program and doing without its benefits

$$WTP_i = t_i^* = \frac{\sum_{k=2}^{K} \beta_k x_{ki} + \varepsilon_i}{\beta_1}$$

Willingness to pay (WTP)

WTP is that program cost which makes the individual just indifferent between

- paying for the program and enjoying its benefits, and
- not paying for the program and doing without its benefits

$$WTP_i = t_i^* = \frac{\sum_{k=2}^{K} \beta_k x_{ki} + \varepsilon_i}{\beta_1}$$

Expected WTP, given *true* parameter values, is the expectation across ε_i, which is a mean-zero error term.

$$E\left[WTP_i\right] = \left[\frac{\sum_{k=2}^{K} \beta_k x_{ki}}{\beta_1}\right] + E\left[\frac{\varepsilon_i}{\beta_1}\right]$$

Benefits and Costs of Attention

Benefits from attention to marginal attribute?

- *Avoided expected utility loss* from wrong choice when attribute is overlooked
- Reflects consequentiality of choice problem

Benefits and Costs of Attention

Benefits from attention to marginal attribute?

- *Avoided expected utility loss* from wrong choice when attribute is overlooked
- Reflects consequentiality of choice problem

Cost of attention to marginal attribute?

- Depends on cognitive abilities
- Depends on time budget (op cost of time)
- Can differ by attribute: order in attribute list, "fine print," "contact dealer for price"

Benefits and Costs of Attention

Benefits from attention to marginal attribute?

- *Avoided expected utility loss* from wrong choice when attribute is overlooked
- Reflects consequentiality of choice problem

Cost of attention to marginal attribute?

- Depends on cognitive abilities
- Depends on time budget (op cost of time)
- Can differ by attribute: order in attribute list, "fine print," "contact dealer for price"

This paper?

- Our data have insufficient variation in costs of attention to different attributes...so treat as constant
- Focus on *benefits* side of story

Expected Utility Loss from Wrong Choice

		Optimal choice (full information)	
		0	1
Observed choice (incomplete information)	0	No lost utility Pr(0 chosen $\cap$ 0 optimal)	
	1		No lost utility Pr(1 chosen $\cap$ 1 optimal)

Expected Utility Loss from Wrong Choice

Observed choice (incomplete information)	Optimal choice (full information): 0	Optimal choice (full information): 1
0	No lost utility Pr(0 chosen $\cap$ 0 optimal)	Utility loss $= V^1 - V^0$ Pr(0 chosen $\cap$ 1 optimal)
1	Utility loss $= V^0 - V^1$ Pr(1 chosen $\cap$ 0 optimal)	No lost utility Pr(1 chosen $\cap$ 1 optimal)

Expected Utility Loss from Wrong Choice

Observed choice (incomplete information)	Optimal choice (full information): beer	Optimal choice (full information): V8
beer	No lost utility Pr(beer chosen $\cap$ beer optimal)	Utility loss $= V^{V8} - V^{beer}$ Pr(beer chosen $\cap$ V8 optimal)
V8	Utility loss $= V^{beer} - V^{V8}$ Pr(V8 chosen $\cap$ beer optimal)	No lost utility Pr(V8 chosen $\cap$ V8 optimal)

Expected Utility Loss from Wrong Choice

		Optimal choice (full information): Beer	Optimal choice (full information): V8
Observed choice (incomplete information)	beer	No lost utility Pr(beer chosen $\cap$ beer optimal)	Utility loss $= V^{V8} - V^{beer}$ Pr(beer chosen $\cap$ V8 optimal)
	V8	Utility loss $= V^{beer} - V^{V8}$ Pr(V8 chosen $\cap$ beer optimal)	No lost utility Pr(V8 chosen $\cap$ V8 optimal)

"I could've had a **V8 !**"

Expected Utility Loss from Wrong Choice

Observed choice (incomplete information)	Optimal choice (full information): Beer	Optimal choice (full information): V8
beer	No lost utility Pr(beer chosen $\cap$ beer optimal)	Utility loss $= V^{V8} - V^{beer}$ Pr(beer chosen $\cap$ V8 optimal)
V8	Utility loss $= V^{beer} - V^{V8}$ Pr(V8 chosen $\cap$ beer optimal)	No lost utility Pr(V8 chosen $\cap$ V8 optimal)

"I could've had a **beer !**"

Expected Utility Loss from Wrong Choice

To calculate expected utility loss from a wrong choice, need:

- The probability of each way this could happen
- The amount of utility lost in each case

Observed choice (incomplete information)	Optimal choice (full information): Beer	Optimal choice (full information): V8
beer	Utility loss = 0 Pr(beer chosen ∩ beer optimal)	Utility loss = $V^{V8} - V^{beer}$ Pr(beer chosen ∩ V8 optimal)
V8	Utility loss = $V^{beer} - V^{V8}$ Pr(V8 chosen ∩ beer optimal)	Utility loss = 0 Pr(V8 chosen ∩ V8 optimal)

Expected Utility Loss from Wrong Choice

Random Utility Model (RUM) error term? ("ε")

- Stuff that is known to the respondent, but unobserved by the investigator
- Assume ε *remains* fully known to the respondent, regardless of whether he/she pays attention to some specific attribute, k, in the choice scenario
- Important: then *same error distribution* is involved, with or without attention to k^{th} specific attibute

Expected Utility Loss from Wrong Choice - binary case

Given that there are two ways to lose utility by a wrong choice when information is ignored:

$$E[U\ Loss] = Pr[1\ chosen|0\ optimal]\ (V_i^0 - V_i^1)$$
$$+Pr[0\ chosen|1\ optimal]\ (V_i^1 - V_i^0)$$

where $V_i^1 - V_i^0 = x_i^{'}\beta + \varepsilon_i$

$$= \underbrace{x_{-ki}^{'}\beta_{-k}}_{\textit{other}\text{-attribs}} + \underbrace{x_{ki}\beta_k}_{\textit{own}\text{-attrib}} + \underbrace{\varepsilon_i}_{\text{error difference}}$$

where $x_{ki} = X_{ki}^1 - X_{ki}^0$ is the kth attribute-difference

Expected Utility Loss from Wrong Choice - binary case

Choice probabilities based upon *complete information*:

- $Pr(1\ optimal) = \Pr\left[x_i^{'}\beta + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_i^{'}\beta\right]$
- $Pr(0\ optimal) = \Pr\left[\varepsilon_i > x_i^{'}\beta\right]$

Expected Utility Loss from Wrong Choice - binary case

Choice probabilities based upon *complete information*:

- $Pr(1\ optimal) = \Pr\left[x_i^{'}\beta + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_i^{'}\beta\right]$
- $Pr(0\ optimal) = \Pr\left[\varepsilon_i > x_i^{'}\beta\right]$

Choice probabilities based on *all but the kth attribute*:

- $Pr(1\ chosen) = \Pr\left[x_{-ki}^{'}\beta_{-k} + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_{-ki}^{'}\beta_{-k}\right]$
- $Pr(0\ chosen) = \Pr\left[\varepsilon_i > x_{-ki}^{'}\beta_{-k}\right]$

Expected Utility Loss from Wrong Choice - binary case

Choice probabilities based upon *complete information*:

- $Pr(1\ optimal) = \Pr\left[x_i^{'}\beta + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_i^{'}\beta\right]$
- $Pr(0\ optimal) = \Pr\left[\varepsilon_i > x_i^{'}\beta\right]$

Choice probabilities based on *all but the kth attribute*:

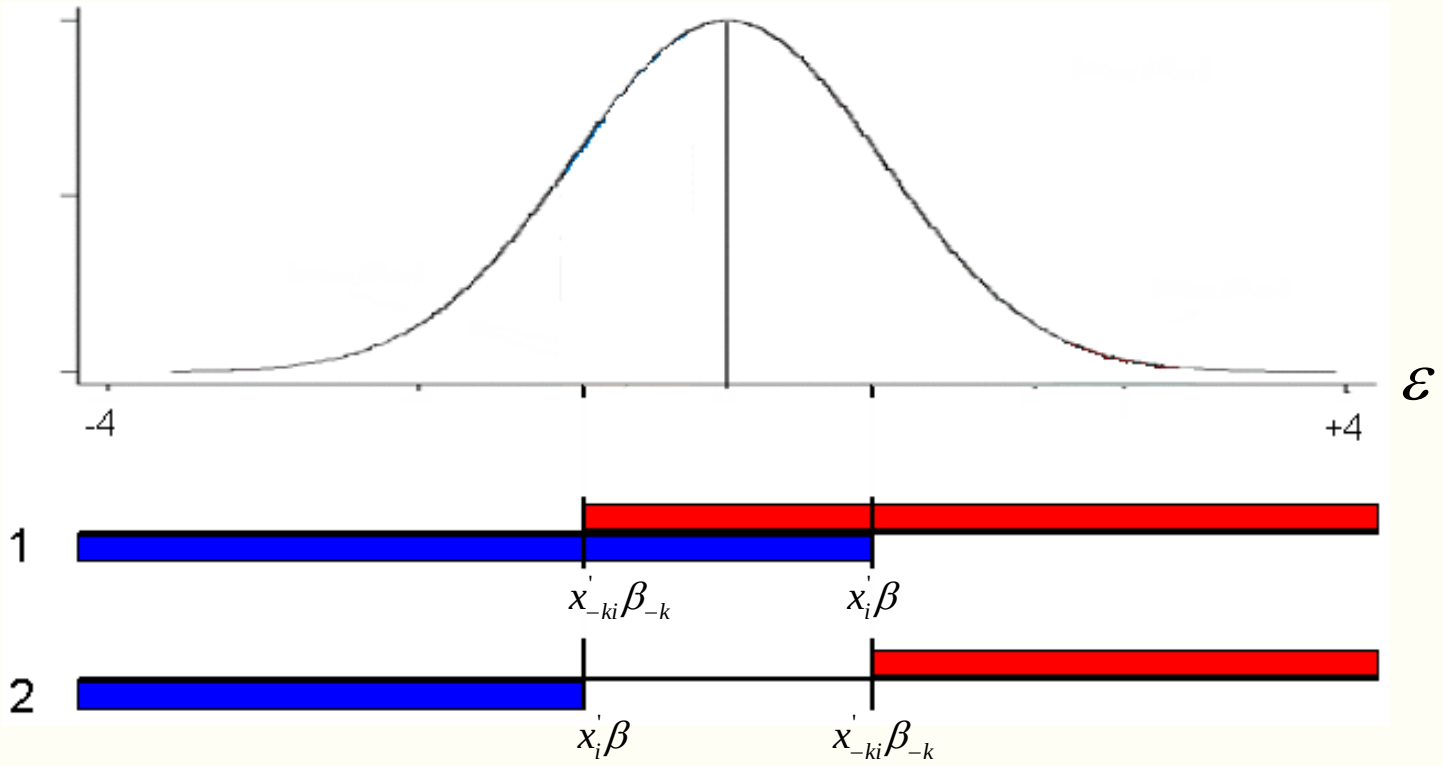
- $Pr(1\ chosen) = \Pr\left[x_{-ki}^{'}\beta_{-k} + \varepsilon_i > 0\right] = \Pr\left[\varepsilon_i < x_{-ki}^{'}\beta_{-k}\right]$
- $Pr(0\ chosen) = \Pr\left[\varepsilon_i > x_{-ki}^{'}\beta_{-k}\right]$

Probability of *wrong choice* when k^{th} attribute is ignored:

$$Pr(1\ optimal \cap 0\ chosen) = \Pr\left[\left(\varepsilon_i < x_i^{'}\beta\right) \cap \left(\varepsilon_i > x_{-ki}^{'}\beta_{-k}\right)\right]$$

$$Pr(0\ optimal \cap 1\ chosen) = \Pr\left[\left(\varepsilon_i > x_i^{'}\beta\right) \cap \left(\varepsilon_i < x_{-ki}^{'}\beta_{-k}\right)\right]$$

Empty versus non-empty sets?



Case 1: if utility from k^th attribute, $x_{ki}\beta_k$, is positive, intervals overlap, probability > 0

Case 2: if utility from k^th attribute, $x_{ki}\beta_k$, is negative, no overlap, probability is zero

Probability of one type of choice mistake

Given that $x'_{-ki}\beta_{-k} + x_{ki}\beta_k = x'_i\beta$, and same ε_i ...

$$\Pr(1\ \textit{optimal} \cap 0\ \textit{chosen})$$
$$= \Pr\left[\varepsilon_i < x'_i\beta \ \cap \ \varepsilon_i > x'_{-ki}\beta_{-k}\right] \quad \text{...substitute, rearrange}$$
$$= \Pr\left[\left(x'_{-ki}\beta_{-k}\right) < \varepsilon_i < \left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right)\right]$$
$$= F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)$$

...can be nonzero only when $x_{ki}\beta_k$ is positive

...i.e. attention to k^{th} attribute would have made alt. 1 look better

...will be differences in cumulative densities over a range given by $x_{ki}\beta_k$, the contribution to utility by k^{th} attribute

Probability of the other type of choice mistake

Given that $x'_{-ki}\beta_{-k} + x_{ki}\beta_k = x'_i\beta$, and same ε_i ...

$$\Pr(0\ \textit{optimal} \cap 1\ \textit{chosen})$$
$$= \Pr\left[\varepsilon_i > x'_i\beta \ \cap \ \varepsilon_i < x'_{-ki}\beta_{-k}\right] \quad \text{...substitute, rearrange}$$
$$= \Pr\left[\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) < \varepsilon_i < \left(x'_{-ki}\beta_{-k}\right)\right]$$
$$= F\left(x'_{-ki}\beta_{-k}\right) - F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right)$$

...can be nonzero only when $x_{ki}\beta_k$ is negative

...i.e. attention to k^{th} attribute would have made alt. 0 look better

...will be differences in cumulative densities over a range given by $x_{ki}\beta_k$, the contribution to utility by k^{th} attribute

Review: Two Ways to Make a Mistake

Observed choice (incomplete information)	Optimal choice (full information): 0	Optimal choice (full information): 1
0	Utility loss $= 0$ Pr(0 chosen $\cap$ 0 optimal)	Utility loss $= V^1 - V^0$ Pr(0 chosen $\cap$ 1 optimal)
1	Utility loss $= V^0 - V^1$ Pr(1 chosen $\cap$ 0 optimal)	Utility loss $= 0$ Pr(1 chosen $\cap$ 1 optimal)

Expected utility loss due to a wrong choice

Expectation is each probability times the associated lost utility:

$$
\begin{aligned}
E[U\ loss] &= \left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]\left(V^1 - V^0\right) \\
&\quad + \left[F\left(x'_{-ki}\beta_{-k}\right) - F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right)\right]\left(V^0 - V^1\right) \\
&= 2\left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]\left(V^1 - V^0\right)
\end{aligned}
$$

Expected utility loss due to a wrong choice

Expectation is each probability times the associated lost utility:

$$E[U\ loss] = \left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]\left(V^1 - V^0\right)$$
$$+ \left[F\left(x'_{-ki}\beta_{-k}\right) - F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right)\right]\left(V^0 - V^1\right)$$
$$= 2\left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]\left(V^1 - V^0\right)$$

Either $\left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]$ and $\left(V^1 - V^0\right)$ are both positive, or they are both negative.

Expected utility loss due to a wrong choice

Expectation is each probability times the associated lost utility:

$$E[U\ loss] = \left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]\left(V^1 - V^0\right)$$
$$+ \left[F\left(x'_{-ki}\beta_{-k}\right) - F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right)\right]\left(V^0 - V^1\right)$$
$$= 2\left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]\left(V^1 - V^0\right)$$

Either $\left[F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right]$ and $\left(V^1 - V^0\right)$ are both positive, or they are both negative.

$$E[U\ loss] = 2\left|F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right| \ \left|x'_i\beta + \varepsilon_i\right|$$

Expected utility loss due to a wrong choice

Benefits from attention to k^{th} attribute increase in the expected utility loss from making a wrong choice by failing to consider this attribute.

This expected utility loss is given by:

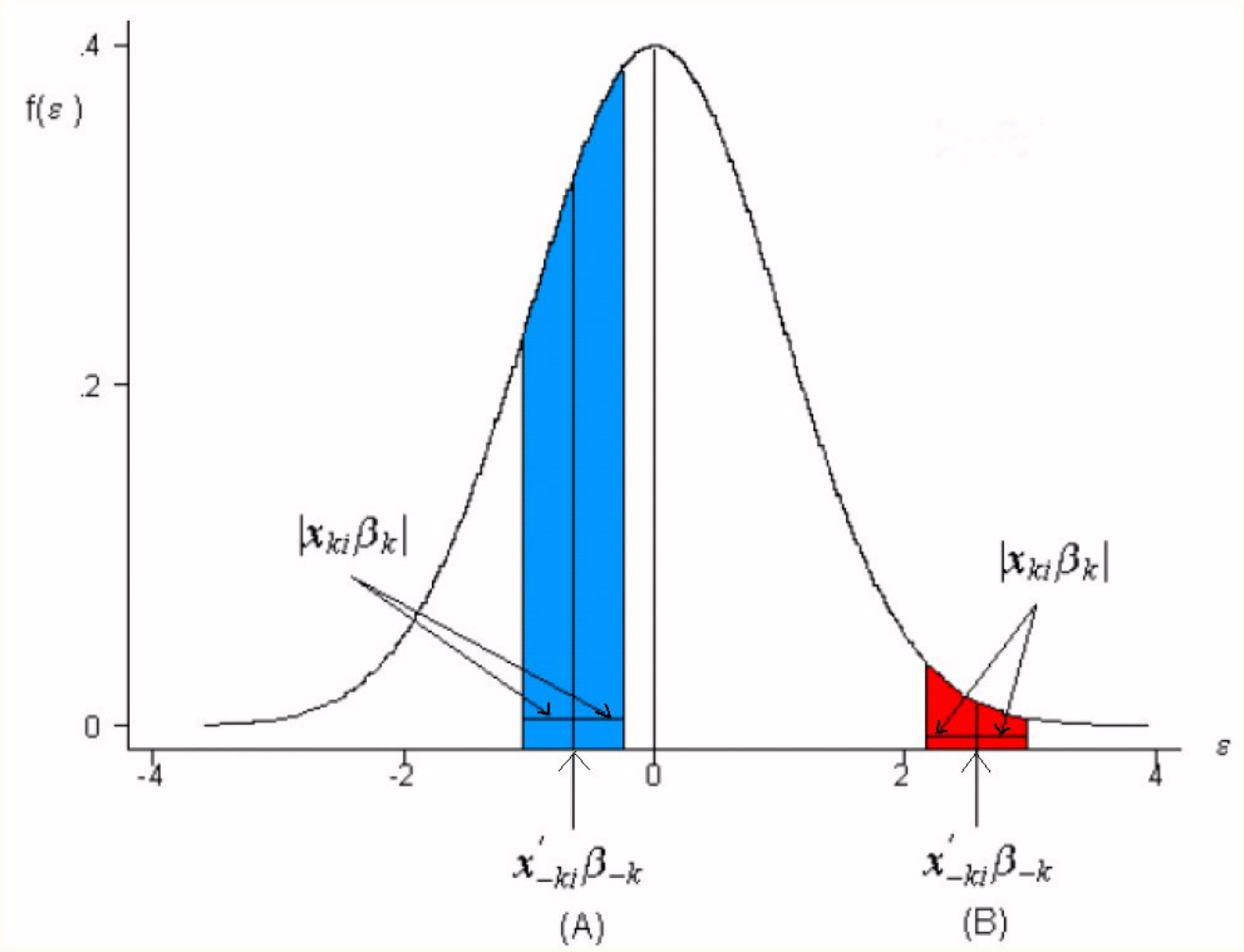
$$E[U\ loss] = 2\left|F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right|\left|x'_i\beta + \varepsilon_i\right|$$

Will be larger as the true but unobserved utility difference, $V^1 - V^0 = x'_i\beta + \varepsilon_i$, is larger in absolute value

For a given *unobserved* utility difference, $E[U\ loss]$ will be larger as more of the probability density for ε is captured within an interval of width $x_{ki}\beta_k$, anchored at $x'_{-ki}\beta_{-k}$

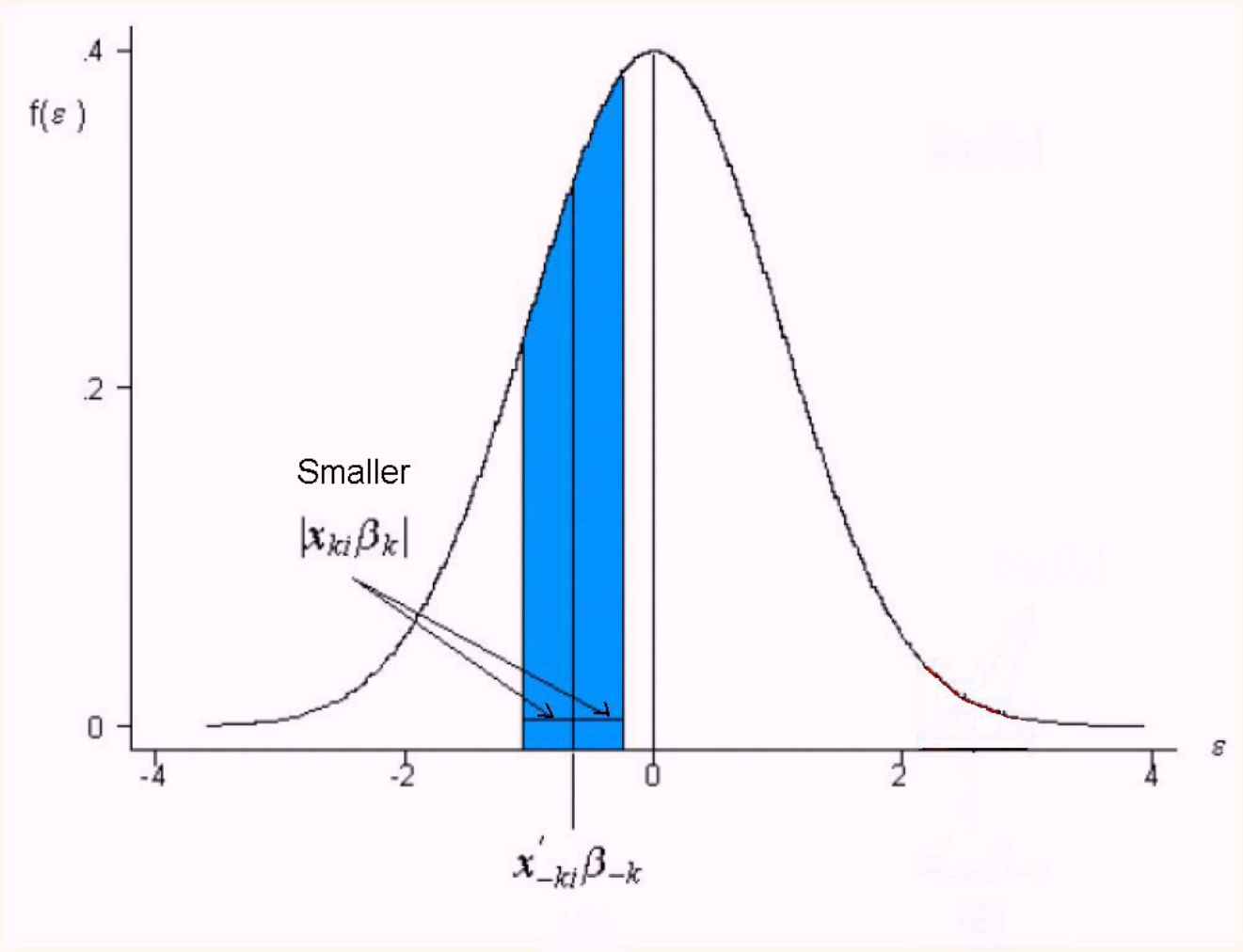
Effect of $\left|x'_{-ki}\beta_{-k}\right|$ on $E[U\ loss]$

An interval of a given width captures more probability if it is near the center of the distribution



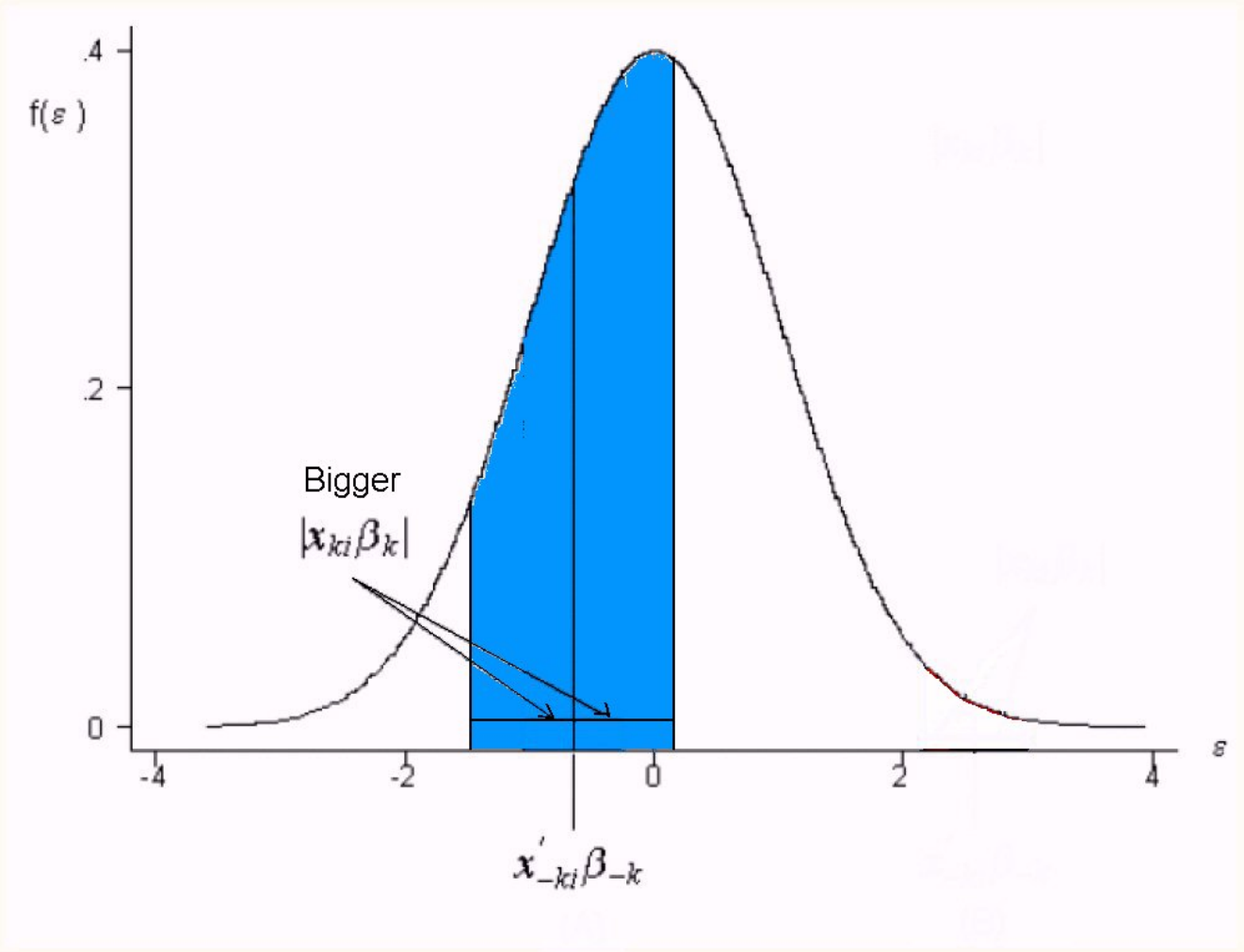
Effect of $|x_{ki}\beta_k|$ on $E[U\ loss]$

For any given anchoring point, a wider interval captures more probability



Effect of $|x_{ki}\beta_k|$ on $E[U\ loss]$

For any given anchoring point, a wider interval captures more probability



Own-attribute utility differences

Interval of ε density: $x_{ki}\beta_k$ = "own-attribute utility difference"

Interval width, $|x_{ki}\beta_k|$ will be larger

- if x_{ki} (amount of attribute) is large in absolute magnitude
- if β_k (its marginal utility) is large in absolute terms

Own-attribute utility differences

Interval of ε density: $x_{ki}\beta_k$ = "own-attribute utility difference"

Interval width, $|x_{ki}\beta_k|$ will be larger

- if x_{ki} (amount of attribute) is large in absolute magnitude
- if β_k (its marginal utility) is large in absolute terms

Implications

- We expect that the propensity to attend to the kth attribute will be greater, the greater the (positive or negative) contribution of any given amount of this attribute to overall utility levels, (β_k)
- If an attribute does not differ at all across alternatives, it should get little attention in the choice process (e.g. $x_{ki} = 0$)

Other-attribute utility differences

$x'_{-ki}\beta_{-k}$ = "other-attribute utility difference"

For a given value of $|x_{ki}\beta_k|$, the absolute difference $\left|F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right|$ will be larger as the amount of cumulative density in this given-width interval of the distribution of ε_i is larger.

Other-attribute utility differences

$x'_{-ki}\beta_{-k}$ = "other-attribute utility difference"

For a given value of $|x_{ki}\beta_k|$, the absolute difference $\left|F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right|$ will be larger as the amount of cumulative density in this given-width interval of the distribution of ε_i is larger.

This captured density is larger:

- when $x'_{-ki}\beta_{-k}$ lies nearer to zero (as opposed to farther out in either tail of the distribution)

Other-attribute utility differences

$x'_{-ki}\beta_{-k}$ = "other-attribute utility difference"

For a given value of $|x_{ki}\beta_k|$, the absolute difference $\left|F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right|$ will be larger as the amount of cumulative density in this given-width interval of the distribution of ε_i is larger.

This captured density is larger:

- when $x'_{-ki}\beta_{-k}$ lies nearer to zero (as opposed to farther out in either tail of the distribution)
- when the indirect utility-difference across alternatives, ignoring the kth attribute, is nearer to zero

Other-attribute utility differences

$x'_{-ki}\beta_{-k}$ = "other-attribute utility difference"

For a given value of $|x_{ki}\beta_k|$, the absolute difference $\left|F\left(x'_{-ki}\beta_{-k} + x_{ki}\beta_k\right) - F\left(x'_{-ki}\beta_{-k}\right)\right|$ will be larger as the amount of cumulative density in this given-width interval of the distribution of ε_i is larger.

This captured density is larger:

- when $x'_{-ki}\beta_{-k}$ lies nearer to zero (as opposed to farther out in either tail of the distribution)
- when the indirect utility-difference across alternatives, ignoring the kth attribute, is nearer to zero
- In words, when the alternatives confer similar utility, in terms of all *other* attributes

Dissimilarity based on other attributes?

With just two alternatives

- The simple absolute difference in systematic utilities according to other attributes, $\left| x'_{-ki}\beta_{-k} \right|$ will adequately capture the relevant properties of the choice set.

Dissimilarity based on other attributes?

With just <u>two</u> alternatives

- The simple absolute difference in systematic utilities according to other attributes, $\left|x'_{-ki}\beta_{-k}\right|$ will adequately capture the relevant properties of the choice set.

With <u>three or more</u> alternatives

- Need to resort to analog measures:

$$dissim(x'_{-ki}\beta_{-k})$$

. . . the extent to which there is a clear-cut "best" option among the available alternatives, based on all attributes *other than* this one.

Candidate measures for $dissim(x'_{-ki}\beta_{-k})$

Candidate 1: $lead(x'_{-ki}\beta_{-k})$

- The utility difference between the two leading alternatives, based on all attributes other than the one in question
- Compute each of the indirect utility differences, relative to the third alternative $x^{1'}_{-ki}\beta_{-k}$, $x^{2'}_{-ki}\beta_{-k}$, and 0
- Identify the maximum and the median values and calculate their absolute difference
- Disadvantage for estimation: not smoothly differentiable

Candidate measures for $dissim(x'_{-ki}\beta_{-k})$

Candidate 2: $stdev(x'_{-ki}\beta_{-k})$

- Standard deviation of $x^{1'}_{-ki}\beta_{-k}$, $x^{2'}_{-ki}\beta_{-k}$, and 0
- The greater the standard deviation in these measures, the more "different" are the alternatives in terms of utility from all other attributes
- Advantage for estimation: differentiable

Candidate measures for $dissim(x'_{-ki}\beta_{-k})$

Candidate 3: $skew(x'_{-ki}\beta_{-k})$

- Skewness of $x^{1'}_{-ki}\beta_{-k}, \quad x^{2'}_{-ki}\beta_{-k}, \quad$ and $\quad 0$
- The more positively skewed, the farther apart are the two highest values, relative to the lowest value
- More of a "clear winner" among the three alternatives in terms of "all but the kth attribute."
- However, can have high skewness but low variance

Candidate 4: entropy measure (e.g. Swait and Adamowicz)

Own-attribute dissimilarity?

With just two alternatives

- The absolute value of the additive component of utility associated with the kth attribute $|x_{ki}\beta_k|$ will adequately capture the relevant term in the theoretical model.

Own-attribute dissimilarity?

With just <u>two</u> alternatives

- The absolute value of the additive component of utility associated with the kth attribute $|x_{ki}\beta_k|$ will adequately capture the relevant term in the theoretical model.

With <u>three or more</u> alternatives

- Need to resort to analog measures:

$$dissim(x_{ki}\beta_k)$$

...the extent to which the utility due to the kth attribute varies "greatly" across the available alternatives.

Analogous candidate measures for $dissim(x_{ki}\beta_k)$

For three alternatives, there will be three terms for the contribution to net utility from the attribute in question:

$$x^1_{ki}\beta_k, \quad x^2_{ki}\beta_k, \quad \text{and} \quad 0$$

Candidate 1: $lead(x_{ki}\beta_k)$

- the absolute size of the "lead" for the largest value

Candidate 2: $stdev(x_{ki}\beta_k)$

- the standard deviation of the three values

Candidate 3: $skew(x_{ki}\beta_k)$

- the skewness of the three values

Data

- We use an existing survey of WTP for health risk reductions by Cameron and DeShazo (2006)
- Survey fielded using the standing consumer panel maintained by Knowledge Networks, Inc.
  - Internet and Web TV
- Use 1519 US subjects
- 79 percent response rate overall (selection effects minimal)
- Pretesting
  - One-on-one focus sessions
  - External panel of distinguished reviewers
  - Canadian sample pre-test (more than 1000)

The Survey

Five Modules:

1. Evaluation of health conditions and perceived threats
2. Illness profile tutorial
   - An illness profile is a sequence of future health states (latency, sick-years, post-illness years and lost life-years)
3. Risk tutorial (Krupnick, Hoffman, et al. 1,000 square risk grid)
4. Conjoint choice sets
   - 3 alternatives per choice set (extensive randomized design) (Program A, Program B, Neither Program)
     - Each program: purchase annual non-invasive test that will reduce probability of illness profile (characterized by illness label, onset, duration, symptoms, treatment and outcome)
   - 5 choice sets per respondent (independent choices)
5. Debriefing questions (studied in Cameron, DeShazo and Johnson, 2007)
   - Allow researcher to know if an individual adjusts the choice scenario and by how much

The Survey: One Randomized Choice Scenario

Choose the program that reduces the illness that you most want to avoid. But think carefully about whether the costs are too high for you. If both programs are too expensive, then choose Neither Program.

If you choose "neither program", remember that you could die early from a number of causes, including the ones described below.

	Program A for Diabetes	**Program B for Heart Attack**
Symptoms/ Treatment	Get sick when 77 years-old 6 weeks of hospitalization No surgery Moderate pain for 7 years	Get sick when 67 years-old No hospitalization No surgery Severe pain for a few hours
Recovery/ Life expectancy	Do not recover Die at 84 instead of 88	Do not recover Die suddenly at 67 instead of 88
Risk Reduction	10% From 10 in 1,000 to 9 in 1,000	10% From 40 in 1,000 to 36 in 1,000
Costs to you	$12 per month [= $144 per year]	$17 per month [= $204 per year]
Your choice	○ **Reduce my chance of diabetes**	○ **Reduce my chance of heart attack**

○ **Neither Program**

Random Utility Model with 3 alternatives

- Simpler version of model in Cameron and DeShazo (2006) (no quadratic or interaction terms, no age heterogeneity)
- Simple linear models are where most choice researchers start
- Choices assumed to maximize present discounted expected utility defined over
  - net income
  - a **complete time profile** of avoided adverse health states, relative to status quo, consisting of illness-years, post-illness years, and lost life-years
- Probabilities? $\Longrightarrow$ expected values ($\Delta\Pi_i^{jS}$ = change Pr(sick))
- Time profiles? $\Longrightarrow$ discounting (5% rate assumed)
- Preliminary work shows utility not linear in present-discounted value (pdv) of health-state years, so use
  - $\log(pdvi + 1)$ for illness-years
  - $\log(pdvr + 1)$ for post-illness (recovered) years
  - $\log(pdvl + 1)$ for lost life-years

Descriptive Statistics

Table 1a: Raw illness/program attributes (14074 programs)

Variable	Description	Mean	Std. Dev.	Min	Max
Cost	Annual cost of program (paid when neither sick nor dead)	355.00	341.14	24	1680
$\Delta\Pi_i^{AS}$	Risk change (i.e. negative, a risk reduction)	-0.0034	0.0017	-0.006	-0.001
Latency	Years until illness/injury begins	19.65	12.03	1	60
Sick years	Duration of illness/injury (years)	6.53	7.21	0	52
Recovered years	Number of years in post-illness health state	1.62	4.62	0	55
Lost life-years	Number of life-years lost	10.87	10.32	0	55

Descriptive Statistics

Table 1b: Constructed attributes (14074 programs)

Variable	Description	Mean	Std. Dev.	Min	Max
(income term)	Net income under each alternative	-0.052747	0.048772	-0.2513	0.1083
$\Delta\Pi_i^{AS}$ *log(pdvi+1)*	Term in present discounted sick-years	-0.003111	0.003006	-0.01710	0
$\Delta\Pi_i^{AS}$ *log(pdvr+1)*	Term in present discounted recovered-years	-0.003374	0.003189	-0.01711	0
$\Delta\Pi_i^{AS}$ *log(pdvl+1)*	Term in present discounted lost life-years	-0.000746	0.001841	-0.01648	0
Sasubrsk (mean = msasubrsk)	Same-illness subjective risk rating (-2 = low, 2=high)	-0.2593	1.2531	-2	2
Cosubrsk (mean = mcosubrsk)	Average subjective risk rating (other major health risks)	-0.2537	0.8670	-2	2
(benefits never)	=1 if expects never to benefit from this program	0.0759	0.2648	0	1
(min overest latency)	Minimum overestimate of the latency of the health risk	-7.483	11.98	-58	29

Descriptive Statistics

Table 1c: Respondent characteristics (1519 respondents)

Variable	Description	Mean	Std. Dev.	Min	Max
Income	Annual income ($)	51,048	33,781	5000	150,000
Female	=1 if female	0.5135	0.5000	0	1
age (mean = mage)	Age in years at time of response	50.11	15.18	25	93

Estimation

Full information maximum likelihood (FIML) estimation is always preferred, but each systematically varying "observed" marginal utility parameter is a function of

- The "true" underlying marginal utility
- Two different (non-trivial) functions of these same parameters (the dissimilarity measures)

Log-likelihood function can be programmed, but optimization is difficult.

Estimation

Resort to an iterative two-step method:

- Estimate initial values of "true" marginal utilities, use to calculate dissimilarity measures

Estimation

Resort to an iterative two-step method:

- Estimate initial values of "true" marginal utilities, use to calculate dissimilarity measures
- Treat dissimilarity measures as exogenous slope-shift variables, estimate new "true" marginal utilities and coefficients on dissimilarity shifters

Estimation

Resort to an iterative two-step method:

- Estimate initial values of "true" marginal utilities, use to calculate dissimilarity measures
- Treat dissimilarity measures as exogenous slope-shift variables, estimate new "true" marginal utilities and coefficients on dissimilarity shifters
- Recalculate dissimilarity measures based on updated "true" marginal utilities

Estimation

Resort to an iterative two-step method:

- Estimate initial values of "true" marginal utilities, use to calculate dissimilarity measures
- Treat dissimilarity measures as exogenous slope-shift variables, estimate new "true" marginal utilities and coefficients on dissimilarity shifters
- Recalculate dissimilarity measures based on updated "true" marginal utilities
- Continue until permutation in parameter vector disappears

Estimation

Resort to an iterative two-step method:

- Estimate initial values of "true" marginal utilities, use to calculate dissimilarity measures
- Treat dissimilarity measures as exogenous slope-shift variables, estimate new "true" marginal utilities and coefficients on dissimilarity shifters
- Recalculate dissimilarity measures based on updated "true" marginal utilities
- Continue until permutation in parameter vector disappears
- Final step standard errors are conditioned on prevailing parameter estimates from last round

Estimation

Resort to an iterative two-step method:

- Estimate initial values of "true" marginal utilities, use to calculate dissimilarity measures
- Treat dissimilarity measures as exogenous slope-shift variables, estimate new "true" marginal utilities and coefficients on dissimilarity shifters
- Recalculate dissimilarity measures based on updated "true" marginal utilities
- Continue until permutation in parameter vector disappears
- Final step standard errors are conditioned on prevailing parameter estimates from last round
- For correct standard errors, insert converged parameters into log-likelihood and calculate numeric Hessian to derive vcov matrix

Estimates

Homogeneous Preferences – Simple linear additive utility

SD1	Income term (β_0)	Sick-years term (α_1)	Recovered-years term (α_2)	Lost life-years term (α_3)
Baseline variable	3.148 (7.77)***	-27.06 (4.50)***	-24.03 (2.51)**	-29.82 (5.68)***
Observations Log L	21111 -10992.674			

Estimates

Heterogeneous Preferences – Several important shifters on marginal utilities

SD2	Income term (β_0)	Sick-years term (α_1)	Recovered-years term (α_2)	Lost life-years term (α_3)
Baseline variable	2.941 (5.16)***	-14.39 (1.93)*	-40.55 (3.98)***	-21.26 (3.23)***
...*female	3.916 (5.58)***	-	-	-
...*(age-mage)	-	-	-1.305 (1.95)*	-
...*(sasubrsk-msasubrsk)	-	-22.09 (3.80)***	-	-40.67 (7.48)***
...*(cosubrsk-mcosubrsk)	-	27.44 (3.20)***	-	30.53 (3.84)***
...*(benefits never)	-	137.4 (4.14)***	-	217.4 (6.63)***
...*(min overest latency)	-	8.13 (12.53)***	-	8.219 (13.65)***
Observations	21111			
Log L	-10326.046			

Estimates

Expected Signs (based on theory)

stdev models	Income term (β_1)	Sick-years term (β_2)	Recovered-years term (β_3)	Lost life-years term (β_4)
Baseline variable	+	-	-	-
...*(sd(U othr attr)-mean sd)	-	+	+	+
...*(sd(U this attr)-mean sd)	+	-	-	-

Estimates ...Hmmm...

Dissimilarity measures based on Homogeneous Preferences specification

SD3; 30 iterations	Income term (β_0)	Sick-years term (α_1)	Recovered-years term (α_2)	Lost life-years term (α_3)
Baseline variable	2.759 (3.96)***	-23.61 (2.46)**	-72.19 (3.69)***	-45.62 (5.18)***
...*(sd(U othr attr)-mean sd)	7.715 (2.86)***	-48.13 (1.04)	-57.4 (0.51)	96.09 (1.83)*
...*(sd(U this attr)-mean sd)	4.882 (1.08)	-4.894 (0.04)	122.3 (1.92)*	102.2 (1.99)**
Observations	21111			
Log L	-10976.589			

Estimates ...Hmmm...

Dissimilarity measures based on Heterogeneous Preferences specification

SD4; 30 iterations	Income term (β_0)	Sick-years term (α_1)	Recovered-years term (α_2)	Lost life-years term (α_3)
Baseline variable	3.455 (5.89)***	-20.57 (2.65)***	-43.81 (2.28)**	-21.59 (3.14)***
...*(sd(U othr attr)-mean sd)	3.152 (3.11)***	-21.01 (1.05)	.3823 (0.01)	17.12 (0.80)
...*(sd(U this attr)-mean sd)	-3.507 (1.92)*	21.08 (0.92)	13.45 (0.13)	-10.95 (0.61)
...*female	5.473 (5.30)***	-	-	-
...*(age-mage)	-	-	-1.348 (1.62)	-
...*(sasubrsk-msasubrsk)	-	-26.32 (4.32)***	-	-41.03 (7.07)***
...*(cosubrsk-mcosubrsk)	-	31.64 (3.61)***	-	30.45 (3.78)***
...*(benefits never)	-	130 (3.82)***	-	215.8 (6.42)***
...*(min overest latency)	-	9.24 (11.88)***	-	8.415 (12.54)***
Observations	21111			
Log L	-10316.015			

Estimates

What might we be missing?

- With homogeneous preferences, the other-attribute utility from any given configuration of attributes will be identical across individuals...insufficient variation?
- What if researcher is estimating a naive linear specification assuming homogeneous preferences, but any given configuration of "other attributes" is viewed differently by different consumers?
- Suppose the heterogeneity in actual preferences creates differences across people in the other-attribute and own-attribute dissimilarity measures.
- How do the researcher's marginal utility estimates from the naive model vary with the level of this "unobservable to the researcher" heterogeneity in the dissimilarity measures?

Estimates ....Ah hah!

Dissimilarity measures based on Heterogeneous Preferences specification

SD5	Income term (β_0)	Sick-years term (α_1)	Recovered-years term (α_2)	Lost life-years term (α_3)
Baseline variable	1.514 (2.93)***	-7.124 (1.02)	-33.92 (2.16)**	-20.23 (3.30)***
...*(sd(U othr attr)-mean sd)	-1.505 (1.78)*	91.18 (5.36)***	31.98 (1.49)	94.09 (4.95)***
...*(sd(U this attr)-mean sd)	2.361 (1.89)*	-106.1 (6.48)***	56.41 (0.68)	-43.3 (3.78)***
Observations	21111			
Log L	-10915.004			

Estimates ....Again

Dissimilarity measures based on Heterogeneous Preferences specification

Lead5; 30 iterations	Income term (β_0)	Sick-years term (α_1)	Recovered-years term (α_2)	Lost life-years term (α_3)
Baseline variable	2.685 (5.82)***	-11.23 (1.69)*	-45.3 (2.99)***	-7.846 (1.33)
...*(ld(U othr attr)-mean ld)	-.4693 (0.74)	63.95 (4.24)***	39.68 (2.12)**	78.25 (4.62)***
...*(ld(U this attr)-mean ld)	1.545 (1.38)	-136.8 (10.94)***	32.27 (0.58)	-116.5 (12.04)***
Observations	21111			
Log L	-10771.54			

Etc. ...

- Similar results for "size-of-lead" measure of dissimilarity, but NOT for skewness or entropy measures
- Corrected standard errors case with "stdev" measure of dissimilarity (Matlab, one-step efficient)
- In progress: Matlab version of model with scaling, rather than translation, of marginal utilities (allows signs to be constrained to be always positive or always negative)
- Spun off: analogous story for attention to alternatives

WTP implications

$$WTP = \left[\frac{1}{\beta_{1i}}\right] \left[\begin{array}{r} \beta_{2i} \left\{ \Delta\Pi_i^j \ \log\left(pdvi_i^j + 1\right) \right\} \\ +\beta_{3i} \left\{ \Delta\Pi_i^j \ \log\left(pdvr_i^j + 1\right) \right\} \\ +\beta_{4i} \left\{ \Delta\Pi_i^j \ \log\left(pdvl_i^j + 1\right) \right\} \end{array} \right]$$

WTP implications

$$WTP = \left[\frac{1}{\beta_{1i}}\right] \left[\begin{array}{c} \beta_{2i} \left\{ \Delta\Pi_i^j \ \log\left(pdvi_i^j + 1\right) \right\} \\ +\beta_{3i} \left\{ \Delta\Pi_i^j \ \log\left(pdvr_i^j + 1\right) \right\} \\ +\beta_{4i} \left\{ \Delta\Pi_i^j \ \log\left(pdvl_i^j + 1\right) \right\} \end{array} \right]$$

Nonlinear in shifted logs of PDV of time in each health state.

WTP will depend upon the **health profile** to be avoided and on the change in its probability of occurring, $\Delta\Pi_i^j$.

- MU(Y) forms the *denominator* (β_{1i})
- Other MUs are in the *numerator* (β_{2i}, β_{3i}, β_{4i})

WTP implications

$$WTP = \left[\frac{1}{\beta_{1i}}\right] \left[\begin{array}{c} \beta_{2i} \left\{ \Delta\Pi_i^j \ \log\left(pdvi_i^j + 1\right) \right\} \\ +\beta_{3i} \left\{ \Delta\Pi_i^j \ \log\left(pdvr_i^j + 1\right) \right\} \\ +\beta_{4i} \left\{ \Delta\Pi_i^j \ \log\left(pdvl_i^j + 1\right) \right\} \end{array} \right]$$

Nonlinear in shifted logs of PDV of time in each health state.

WTP will depend upon the **health profile** to be avoided and on the change in its probability of occurring, $\Delta\Pi_i^j$.

- MU(Y) forms the *denominator* (β_{1i})
- Other MUs are in the *numerator* (β_{2i}, β_{3i}, β_{4i})

Need to consider how the two types of dissimilarity variables influence the apparent size of each systematically varying β parameter.

Effects on Marginal Utilities

Apparent marginal utility of kth attribute is

$$\beta_k^* = \beta_k^{true} + \beta_k^{other}(sd(U\ other\ attr) - mean\ sd)) + \beta_k^{own}(sd(U\ own\ attr) - mean\ sd))$$

Effects on Marginal Utilities

Apparent marginal utility of kth attribute is

$$\beta_k^* = \beta_k^{true} + \beta_k^{other}(sd(U\ other\ attr) - mean\ sd)) + \beta_k^{own}(sd(U\ own\ attr) - mean\ sd))$$

Next slide: magnitudes other-attribute and own-attribute dissimilarity effects on apparent MUs

- β_k^{true} (when attention is "leveled "at sample means)
- $\beta_k^* = \beta_k^{true} + \beta_k^{other}(sd(U\ other\ attr) - mean\ sd))$
  - at selected percentiles of the distr. of the shift variable
- $\beta_k^* = \beta_k^{true} + \beta_k^{own}(sd(U\ own\ attr) - mean\ sd))$
  - at selected percentiles of the distr. of the shift variable

Effects on Marginal Utilities

Dissimilarity variables normalized so that sample mean = 0	Denominator of WTP ↓ Income term (β_1)	In numerator of WTP: Sick-years term (β_2)	Recovered-years term (β_3)	Lost life-years term (β_4)
MU at "mean" dissimilarity =	1.514	-7.124	-33.92	-20.23
Effects of other-attribute utility dissimilarity (percentiles):				
5th	2.24	-38.04	-49.38	-47.83
25th	2.01	-28.85	-44.52	-38.91
50th	1.68	-16.37	-37.64	-27.48
75th	1.19	6.03	-27.63	-8.16
95th	0.21	50.81	-4.99	30.20
Effects of own-attribute utility dissimilarity (percentiles):				
5th	1.01	20.08	-35.90	-7.53
25th	1.12	12.77	-35.90	-10.76
50th	1.33	1.08	-35.90	-15.94
75th	1.70	-18.80	-33.47	-25.07
95th	2.67	-62.93	-25.82	-48.23

Conclusions and Caveats

Don't ignore cross-sectional heterogeneity in preferences.

Conclusions and Caveats

Don't ignore cross-sectional heterogeneity in preferences.

At least in *naive* models of choice, preference heterogeneity (interacting with the mix of attributes across alternatives in a choice set) can influence attention to different attributes. For any given attribute, the *apparent* size of MUs can vary directly with

- the extent to which utility-based-on-other-attributes fails to produce a clear winner among alternatives
- the extent to which utility is very different, across alternatives, based on this attribute

Thus, WTP can vary with these factors as well.

Conclusions and Caveats

Suggestion from my group at the Choice Symposium at Wharton:

Marianne Bertrand and Sendhil Mullainathan (AER, 2004) "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination"

Q: Could we design "paired resume" experiments to maximize and/or minimize the apparent effects of race on estimates of an employer's preferences across prospective employees?

A: If candidates are *similar* on all other dimensions, or even if they are just *close substitutes* on all other dimensions, the apparent influence of inferred race on call-back decisions could be exaggerated.

Conclusions and Caveats

Implication: It may be possible (inadvertently or intentionally) to produce different estimates of MUs and thus WTP by steering respondents' relative attention to different attributes via the mixes of attributes presented in their choice sets.

Conclusions and Caveats

Implication: It may be possible (inadvertently or intentionally) to produce different estimates of MUs and thus WTP by steering respondents' relative attention to different attributes via the mixes of attributes presented in their choice sets.

Under incomplete attention, if the cost attribute captures respondents' attention more than other attributes, then we expect inattention to produce a downward bias in WTP estimates.

Conclusions and Caveats

Implication: It may be possible (inadvertently or intentionally) to produce different estimates of MUs and thus WTP by steering respondents' relative attention to different attributes via the mixes of attributes presented in their choice sets.

Under incomplete attention, if the cost attribute captures respondents' attention more than other attributes, then we expect inattention to produce a downward bias in WTP estimates.

Escape hatch? We may be able to avoid attention-based biases to a considerable extent by recognizing systematic heterogeneity in preferences and taking care to accommodate it in our empirical models.

End of presentation

cameron@uoregon.edu