Measuring Unobservables in Behavioral Models

Joan Walker
Boston University

Swiss Transport Research Conference
September 2007

Outline

- Research agenda
- Methodology overview
- Applications
  1. Residential location
  2. Airline itineraries
  3. Environmentalism
- Conclusion

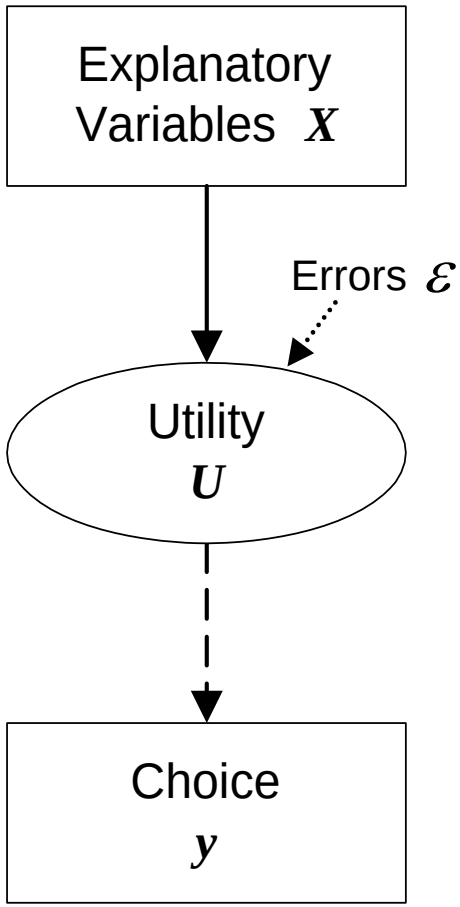
Research Agenda

- Incorporate latent variables and psychometric measurements into discrete choice models
  - Quantify attitudes, motivation, lifestyle, etc.
  - “Behavioral mixing”
- Approach: Integrate discrete choice models and latent variable models

This presentation

- Overview
- 3 Applications
  - Demonstrate that behavioral mixing produces more intuitively appealing models
  - Leading to improved prediction and policy analysis

Discrete Choice Model



Random Utility

$$U = \beta' X + \varepsilon$$

structural equation

Choice Indicator

(Utility Maximization)

$$y = f(U)$$

measurement equation

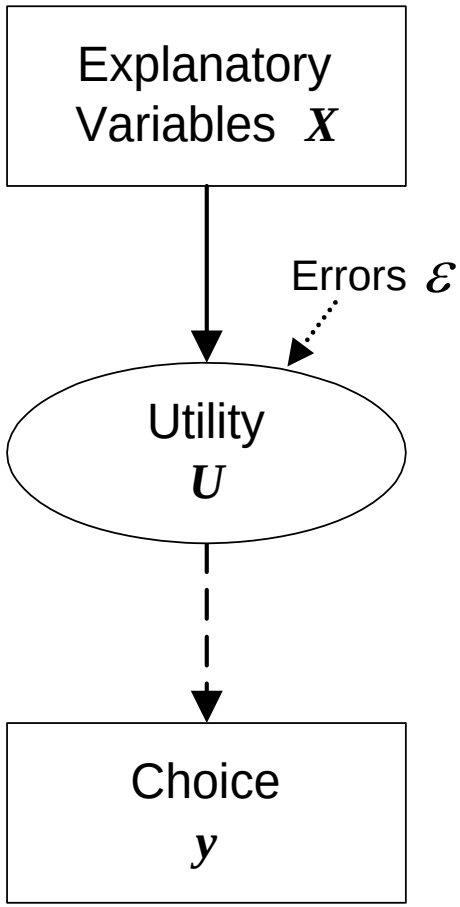
Choice Probability

- Distribution of ε
  - Logit (iid Extreme Value)

$$\mathrm{P}(i \mid X) = \frac{e^{\beta' X_i}}{\sum_{j \in C} e^{\beta' X_j}}$$

  - Probit (Normal)
  - Mixtures (Random β)

Discrete Choice Model



Random Utility

$$U = \beta' X + \varepsilon$$

structural equation

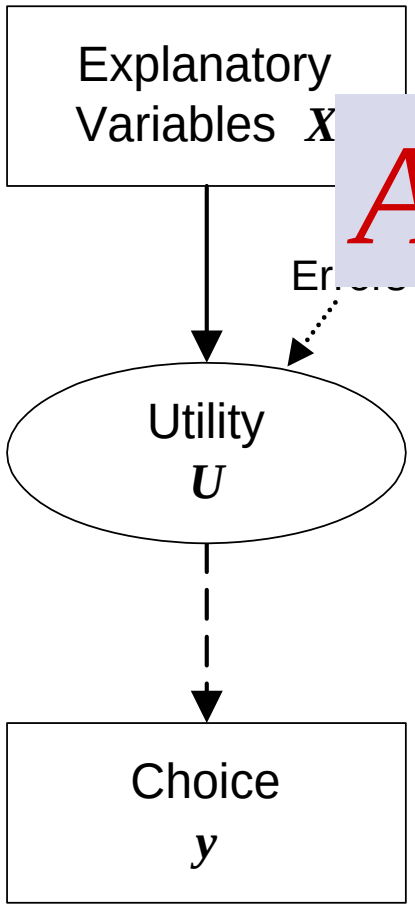
Choice Indicator

(Utility Maximization)

$$y = f(U)$$

measurement equation

Discrete Choice Model



Attitudes?

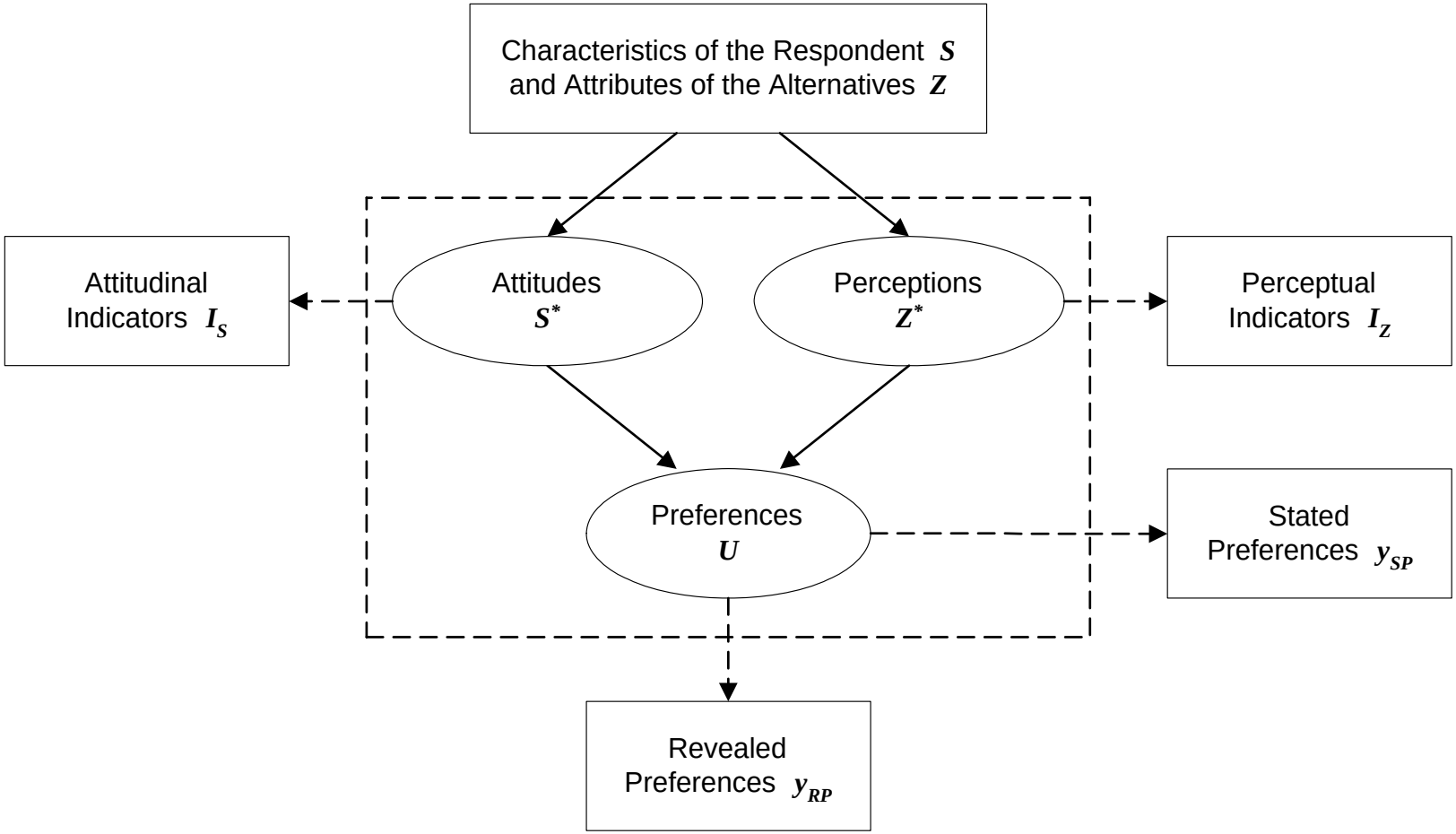
Motivations?

Lifestyles?

$y = f(U)$

tural
ation
measurement equation

Opening the Black Box



APPLICATION I

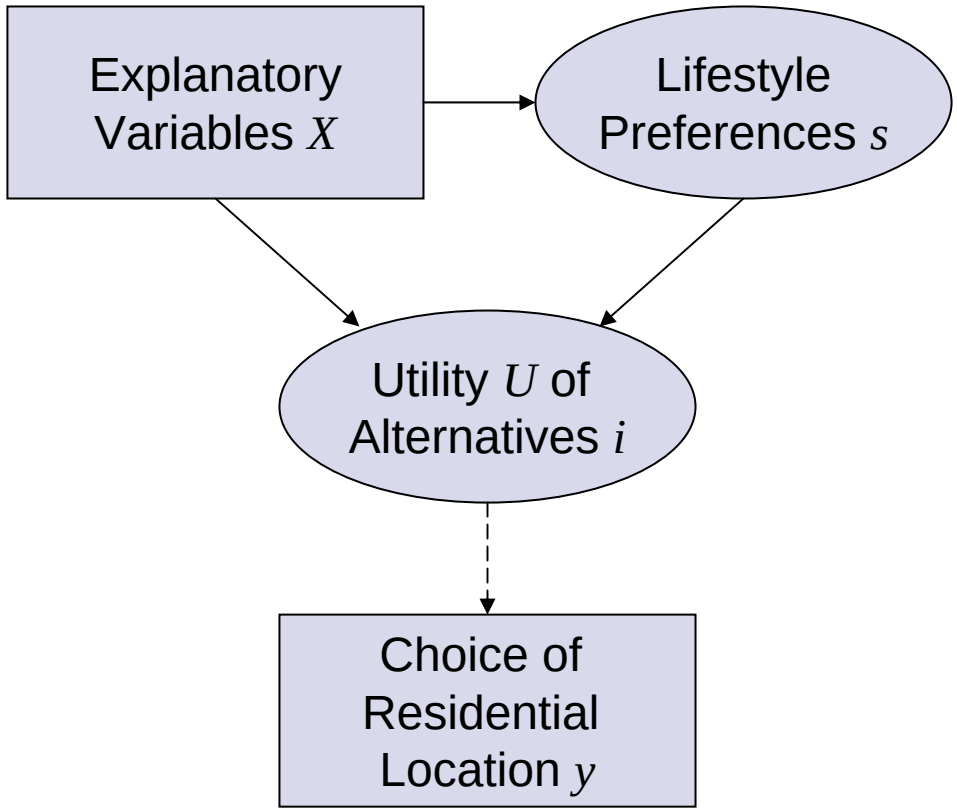
Residential Location & Lifestyle Segmentation

(with Li)

Approach

- Objective: Introduce “lifestyle” in models
- Hypothesis
  - Lifestyle preferences exist
  - Lifestyle differences lead to differences in considerations, criterion, and preferences for residential location choices.
- Infer “lifestyle” preferences from choice behavior using latent class choice model.
  - Latent classes = lifestyle
  - Choice model = location decisions

Residential Location Choice Model with Latent Lifestyle Segmentation



Choice Experiment Example

	(Alternative 1)	(Alternative 2)	(Alternative 3)	(Alternative 4)	(Alt. 5)
	Buy Single Family	Buy Multi-Family	Rent Single Family	Rent Multi-Family	
Type of Dwelling :	*single house*	*apartment*	*duplex / row house*	*condominium*	
Residence Size :	*< 1,000 sq. ft.*	*500-1,000 sq. ft.*	*1,500 - 2,000 sq. ft.*	*< 500 sq. ft.*	Move
Lot Size :	*< 5,000 sq. ft.*	*n/a*	*5,000 - 7,500 sq. ft.*	*n/a*	out
Parking :	*street parking only*	*street parking only*	*driveway, no garage*	*reserved, uncovered*	of the
Price or Monthly Rents :	*< $75K*	*$50K - $100K*	*> $1,200*	*$300 - $600*	Metro
Community Type :	*mixed use*	*mixed use*	*rural*	*urban*	Area
Housing Mix :	*mostly single family*	*mostly multi-family*	*mostly multi-family*	*mostly multi-family*	
Age of Development :	*10-15 years*	*0-5 years*	*10-15 years*	*0 - 5 years*	
Mix of Residential Ownership :	*mostly own*	*mostly own*	*mostly rent*	*mostly own*	
Shops/Services/Entertainment :	*community square*	*basic shops*	*community square*	*basic, specialty shops*	
Local Parks :	*none*	*yes*	*none*	*none*	
Bicycle Paths :	*none*	*yes*	*yes*	*yes*	
School Quality :	*very good*	*very good*	*fair*	*fair*	
Neighborhood Safety :	*average*	*average*	*average*	*average*	
Shopping Prices Relative to Avg :	*20% more*	*20% more*	*same*	*10% more*	
Walking Time to Shops :	*20-30 minutes*	*20-30 minutes*	*< 10 minutes*	*10 - 20 minutes*	
Bus Fare, Travel Time to Shops :	*$1.00, 15-20 minutes*	*$1.00, > 20 minutes*	*$0.50, 5 - 10 minutes*	*$0.50, < 5 minutes*	
Travel Time to Work by Auto :	*> 20 minutes*	*15-20 minutes*	*15 - 20 minutes*	*< 10 minutes*	
Travel Time to Work by Transit :	*> 45 minutes*	*30-45 minutes*	*30 - 45 minutes*	*15 - 30 minutes*	

Residential Location Choice Model with Latent Lifestyle Segmentation

- Location choice model conditional on lifestyle preferences $P_n(i \mid X_n, s)$

  Prob (Location i | "Suburban Lifestyle", other explanatory variables)

- Model of lifestyle preferences $P_n(s \mid X_n)$

  Prob ("Suburban Lifestyle" | Income, Age, etc.)

- Joint estimation

$$P_n(i \mid X_n) = \sum_{s=1}^{S} P_n(i \mid X_n, s) P_n(s \mid X_n)$$

Extension 1: Multiple Responses per Person

- Single response

$$P_n(i \mid X_n) = \sum_{s=1}^{S} P_n(i \mid X_n, s) P_n(s \mid X_n)$$

$$= \sum_{s=1}^{S} P_n(s \mid X_n) P_n(i \mid X_n, s)$$

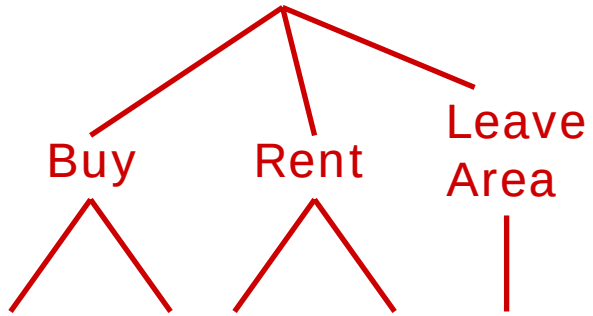
- Multiple responses

$$P_n(i_1, \ldots, i_T \mid X_n) = \sum_{s=1}^{S} P_n(s \mid X_n) \prod_{t=1}^{T} P_n(i_t \mid X_{nt}, s)$$

Extension 2: Continuous Mixing for Nesting

- Error Components

$$
\begin{aligned}
U_{nts}^{BSF} &= \beta_s X_{nt}^{BSF} + \sigma^B \eta_n^B + \varepsilon_{nts}^{BSF} \\
U_{nts}^{BMF} &= \beta_s X_{nt}^{BMF} + \sigma^B \eta_n^B + \varepsilon_{nts}^{BMF} \\
U_{nts}^{RSF} &= \beta_s X_{nt}^{RSF} + \sigma^R \eta_n^R + \varepsilon_{nts}^{RSF} \\
U_{nts}^{RMF} &= \beta_s X_{nt}^{RMF} + \sigma^R \eta_n^R + \varepsilon_{nts}^{RMF} \\
U_{nts}^{LEAVE} &= \beta_s X_{nt}^{LEAVE} + \sigma^L \eta_n^L + \varepsilon_{nts}^{LEAVE}
\end{aligned}
$$



$$\eta_n^B, \eta_n^R, \eta_n^L \sim iid\ N(0,1) \qquad \varepsilon_{nts}^{BSF}, \ldots, \varepsilon_{nts}^{LEAVE} \sim iid\ Extreme\ Value$$

- Choice Model

$$P_n(i_1, \ldots, i_T \mid X_n) = \int \left[\sum_{s=1}^{S} P_n(s \mid X_n) \prod_{t=1}^{T} P_n(i_t \mid X_{nt}, s, \eta) \right] f(\eta) d\eta$$

Estimated with Latent Gold Choice by Statistical Innovations

Overview of Estimation Results

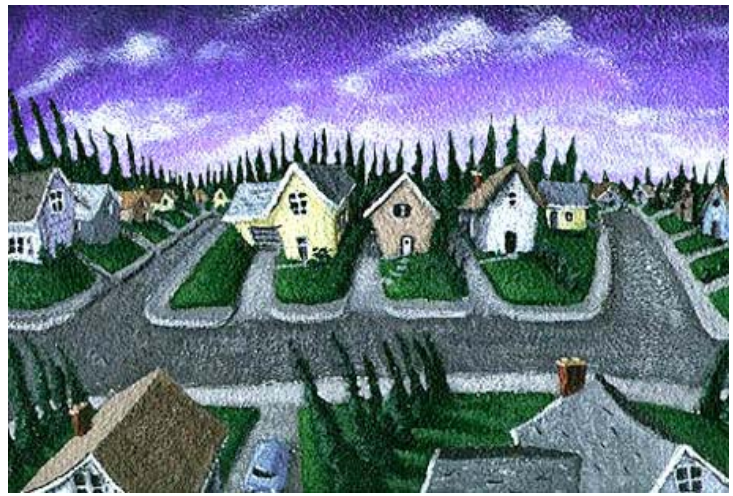
	WITHOUT Lifestyle Segmentation	WITH Lifestyle Segmentation		
Number of classes	1	2	3	4
Number of parameters	37	76	115	155
Rho-bar-square BIC	0.210	**0.213**	0.211	0.204
Rho-bar-square AIC	0.222	0.238	0.248	**0.254**

Chosen Model

Lifestyle Segmentation Results

Latent Segment 1 →

suburban, school, auto… (shopping)
affluent, more established families



← Latent Segment 2

transit, school… (suburban)
less affluent, younger families

Latent Segment 3 →

high density, urban activity… (car)
older, non-family, professionals

“Have it All” mentalities

- Class 1
  - large homes and auto-oriented
  - local high-end shopping
- Class 2
  - suburban lifestyle
  - convenience of transit for work
- Class 3
  - auto-oriented
  - urban
- Relationship with development styles
  - Suburban? Mixed use? Transit-oriented?

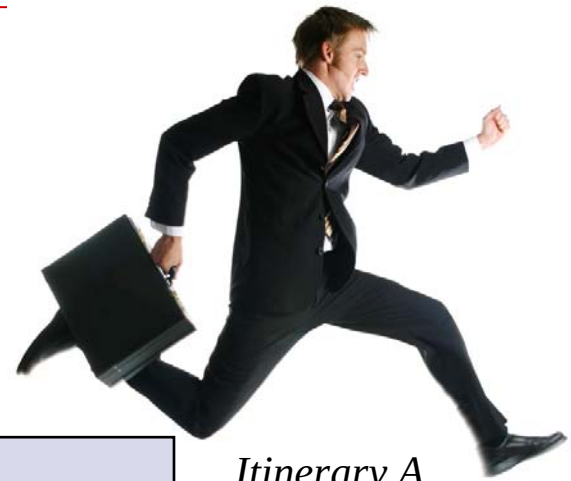
Policy Implications

- How to encouraging more environmentally-positive behavior?
- Design for having it all
  - Supply demand for low-density residential, but support with viable transit to work. (Park-n-ride)
  - Accommodate car in urban (making the lifestyle more appealing to people who otherwise would reject) while making it easier not to have to use the car.

APPLICATION II

Airline Operations

(with Theis)



1:15

Geneva - Frankfurt → *Frankfurt - Boston* — *Itinerary A*

Geneva - Frankfurt → *Frankfurt - Boston* — *Itinerary B*

2:15

Attitudes in Airline Itinerary Choice

- Hypothesis
  - Travelers may choose longer connection time
  - Attitudes: RISK, RUSH, TRUST

- Motivated by airline operating assumptions
  - Demand: added minutes decreases market share
  - Supply: Depeaking lowers supply-side costs

2005 Survey of U.S. Domestic Passengers

- Airline
- Aircraft Type
- Departure Airport
- Flight Times
- **Min. Connecting Time**
- **Buffer Time**
- Number of Connections
- On-time Performance
- Round Trip Fare

Which would you choose for a trip to Jacksonville, FL?

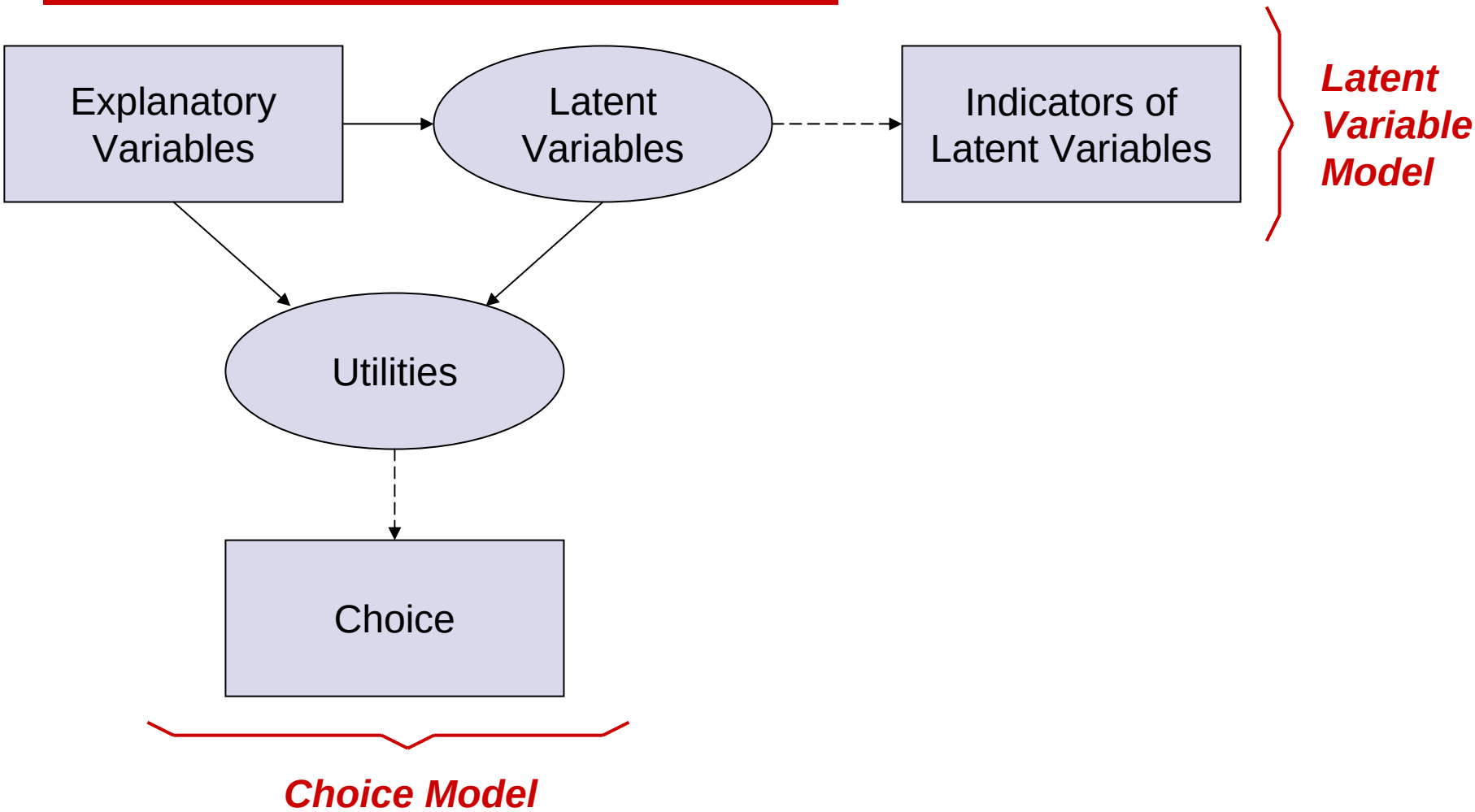
		Your Current Flight	Alternate Flight
AIRLINE		Delta	Continental
AIRCRAFT TYPE		Regional Jet	Standard Jet
DEPARTURE	AIRPORT	Logan International Airport, Boston MA	Burlington International Airport, Burlington VT
	TIME	8:00 AM	5:00 PM
ARRIVAL	AIRPORT	Jacksonville International	Jacksonville International
	TIME	12:00 PM	10:00 PM
LAYOVER TIME		1 hr. (your connecting airport requires a minimum of 40 mins. to connect)	40 mins. (the connecting airport requires a minimum of 40 mins. to connect)
TOTAL TRAVEL TIME		4 hrs.	5 hrs.
NUMBER OF CONNECTIONS		1	1
ON-TIME PERFORMANCE		80% of these flights are on time	90% of these flights are on time
ROUND TRIP FARE		$250	$188
I would choose:		○ my current flight	○ the alternate flight

Psychometric Indicators

- 14 statements rated on a 5 point scale "strongly disagree" to "strongly agree"

I like to take my time when connecting between flights
It's hard for me to find my way through airports
Given two itineraries that only differ in connecting time, I always choose the one with shorter connecting time
I don't think time at airports is wasted because I can shop, eat, or work at airports
I'm willing to accept the risk of a missed connection if this gets me to my destination earlier most of the time
I usually arrive at the check-in counter just before the check-in deadline
Airlines sometimes underestimate the time needed to connect between flights
It is the passenger's responsibility to plan for a sufficient transfer time when booking a connecting itinerary
I don't mind being rushed at a connecting airport if this means I'll arrive at my final destination earlier
Airlines only sell connections that they expect passengers could make
I try to avoid short connections because of the risk of either me or my luggage missing the connecting flight
I enjoy time having extra time at airports
I make sure that the planned connecting time is adequate for me when booking a connecting itinerary

Using Psychometric Indicators



Formulation

- Standard choice model (no latent variables)

$$Prob(y|X;\beta)$$

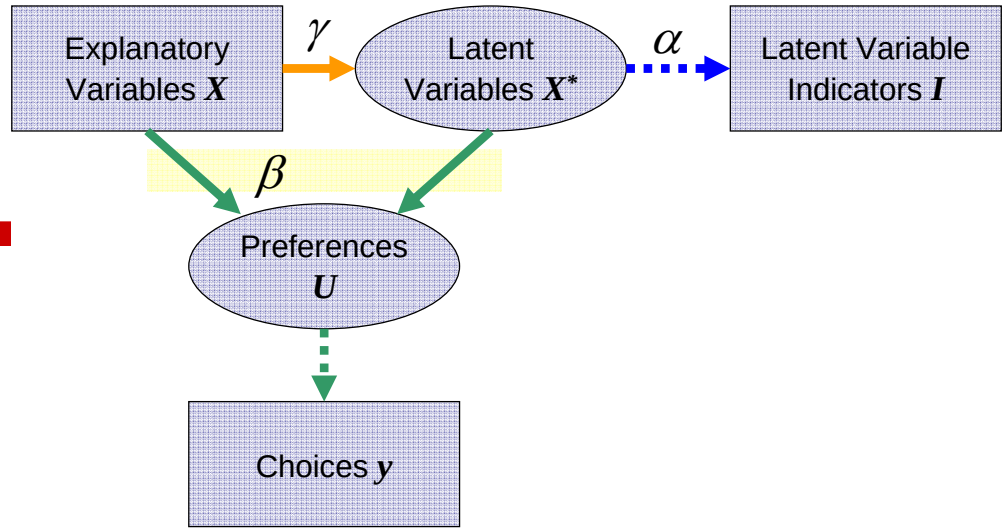
- Choice model with latent variables

$$Prob(y|X,X^*;\beta)$$

- The latent variables are unknown

$$Prob(y|X;\beta,\gamma) = \int_{X^*} Prob(y|X,X^*;\beta)\, f(X^*|X;\gamma)\, dX^*$$

Estimation with Indicators



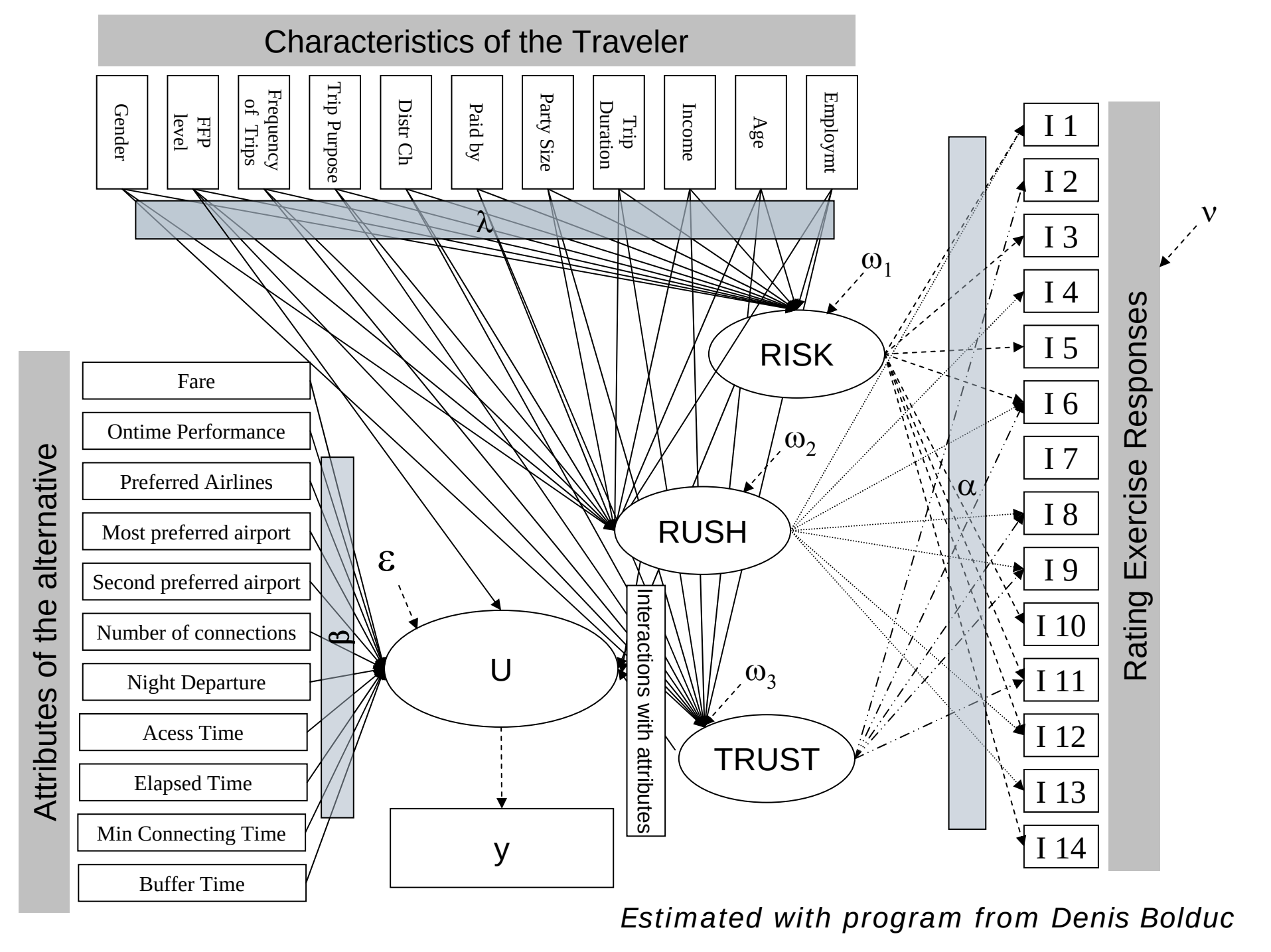
$$f(y, I \mid X; \alpha, \beta, \gamma)$$

$$= \int_{X^*} P(y \mid X, X^*; \beta) f(I \mid X^*; \alpha) f(X^* \mid X; \gamma) dX^*$$

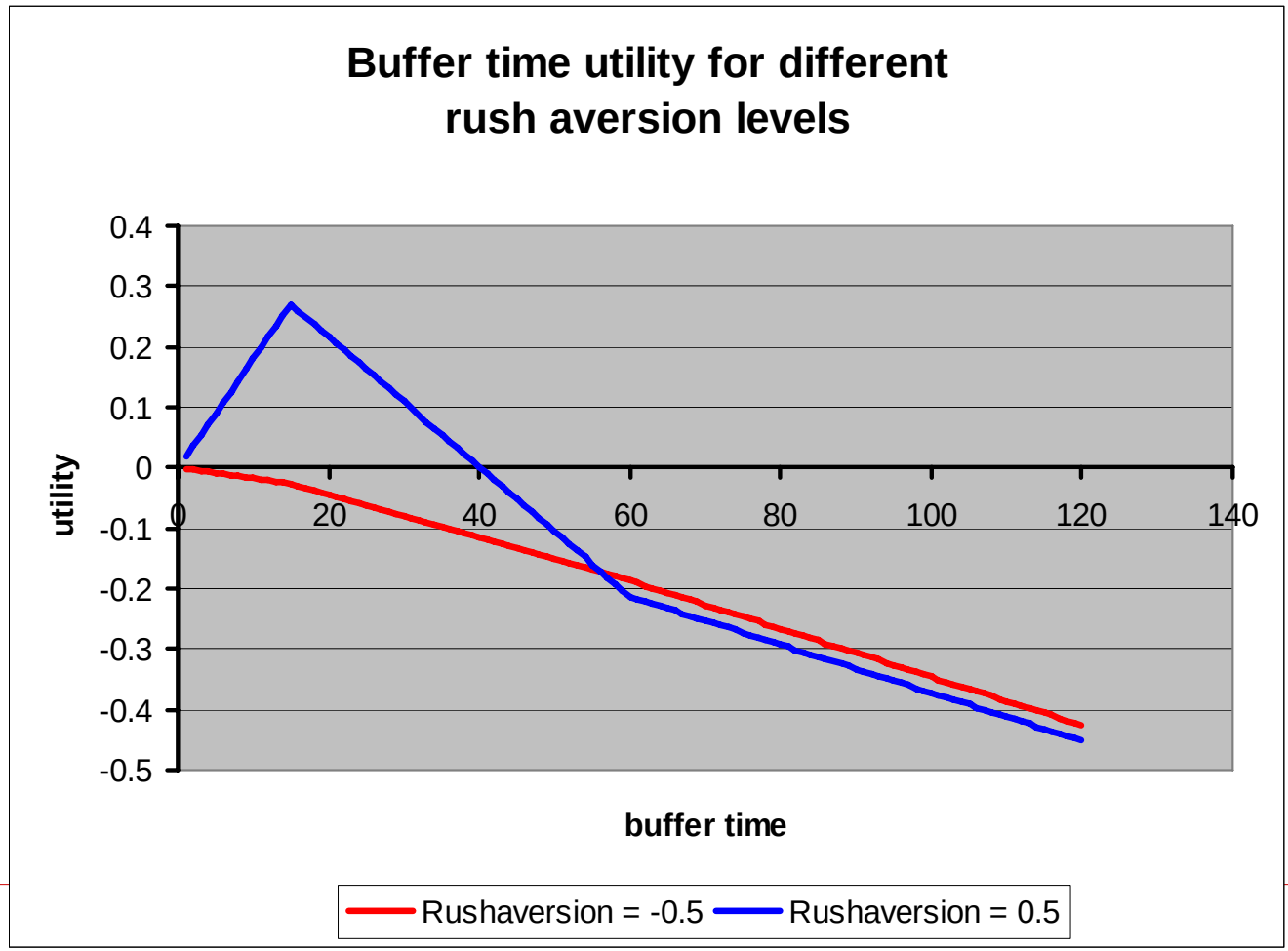
Choice Model

Latent Variable Measurement Model

Latent Variable Structural Model


Characteristics of the Traveler
Gender
FFP level
Frequency of Trips
Trip Purpose
Distr Ch
Paid by
Party Size
Trip Duration
Income
Age
Employmt
λ
ω1
RISK
ω2
RUSH
ω3
TRUST
Interactions with attributes
Attributes of the alternative
Fare
Ontime Performance
Preferred Airlines
Most preferred airport
Second preferred airport
Number of connections
Night Departure
Acess Time
Elapsed Time
Min Connecting Time
Buffer Time
β
ε
U
y
α
ν
Rating Exercise Responses
I 1
I 2
I 3
I 4
I 5
I 6
I 7
I 8
I 9
I 10
I 11
I 12
I 13
I 14
Estimated with program from Denis Bolduc

Influence of Rush Aversion on Perception of Buffer Time



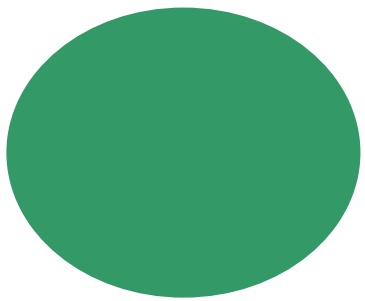
Who Tends Towards Rush Aversion?

- Demographics
  - Females
  - Low/middle income
  - Employed persons
- Trip characteristics
  - Business travelers
  - Checked bags
  - Short trips (< 3 nights)
- Traveling history
  - Not missed flight in last 6 months
  - Not elite travelers

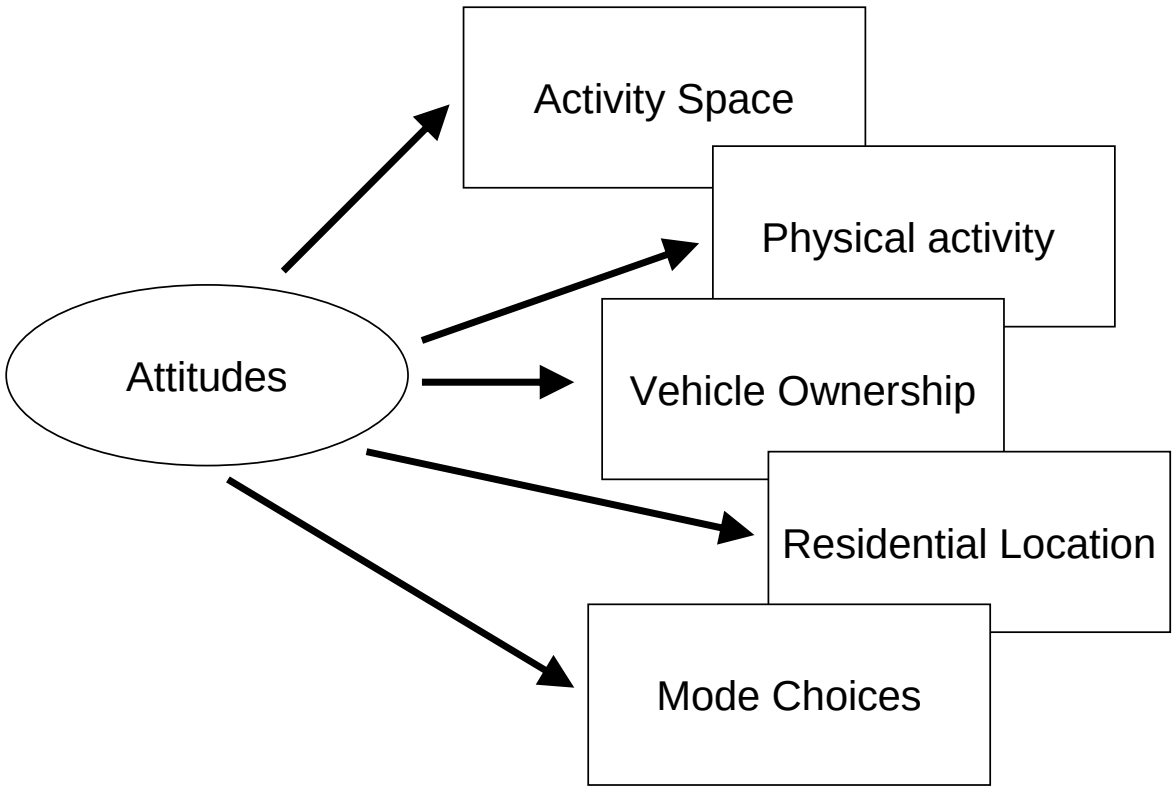
APPLICATION III

Environmentalism and Behavior

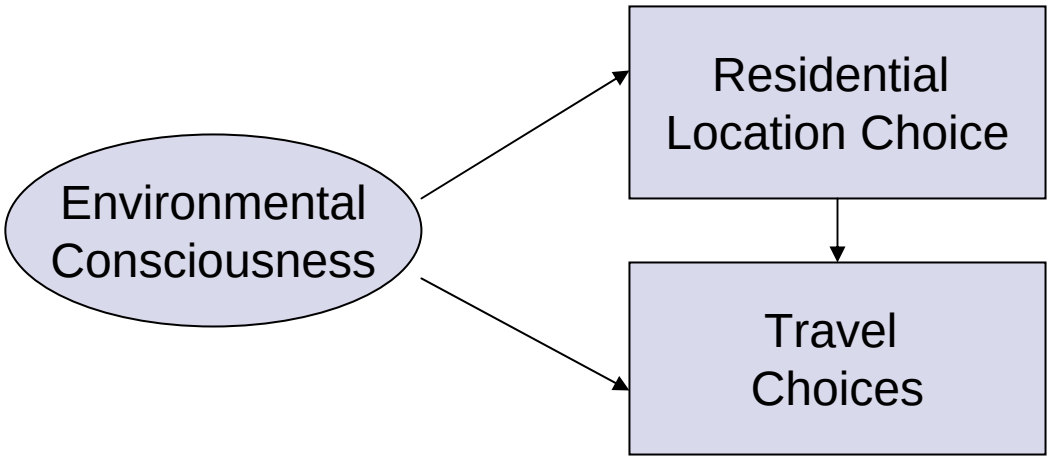
(with Mokhtarian and Schwanen)



Hypothesis: Intrinsic Attitudes Influencing Related Behaviors



Application



- 1993 household travel and activity survey from San Francisco Bay Area[1]
- Multiple neighborhoods sampled
  - "Traditional" – North San Francisco
  - "Suburban" – Concord & San Jose

[1]*Sponsored by California Air Resources Board.*
Described in Kitamura, Laidet, Mokhtarian, Buckinger, Gianelli (1994).
Using data processed by Michael Bagley.

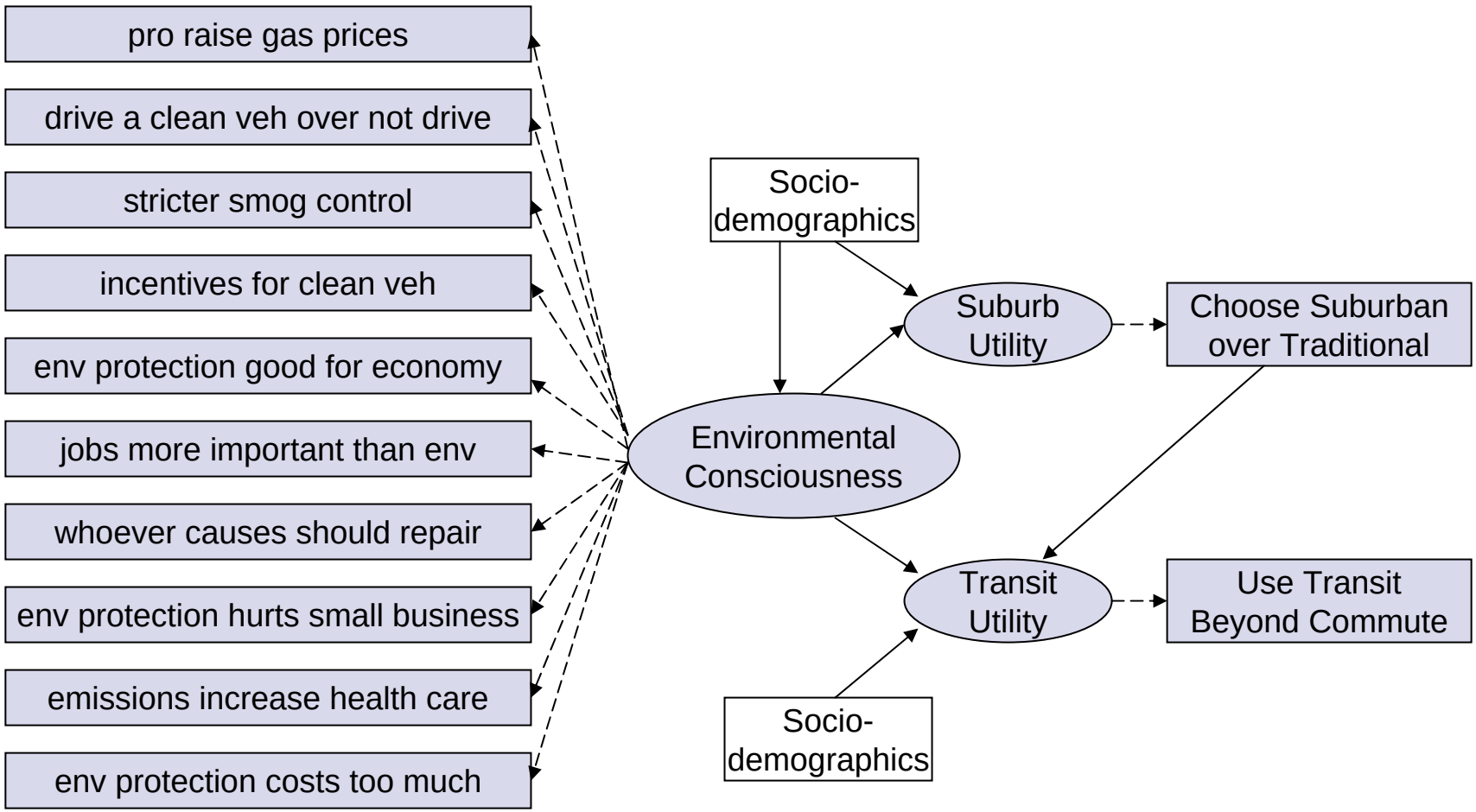
Choices

- Residential Location
  - Suburban Neighborhood
  - Traditional Neighborhood
- Travel
  - Use of transit beyond commuting
  - No use of transit beyond commuting

Indicators of Environmental Consciousness

- We should raise the price of gasoline to reduce congestion and air pollution.
- I would rather drive an electric or other clean-fuel vehicle than give up driving.
- Stricter vehicle smog control laws should be introduced and enforced.
- We should provide incentives to people who use electric or other clean-fuel vehicles
- Environmental protection is good for California's economy.
- People and jobs are more important than the environment.
- Whoever causes environmental damage should repair the damage.
- Environmentalism hurts minority and small businesses.
- Vehicle emissions increase the need for health care.
- Environmental protection costs too much.

Model Framework



Formulation

- Residential location – binary logit

$$U_{Suburb} = \beta_0 + \beta_1 \begin{pmatrix} Environmental \\ Consciousness \end{pmatrix} + \beta_X \begin{pmatrix} Socio- \\ Demographics \end{pmatrix} + \varepsilon_{Suburb}$$

$$U_{Traditional} = \varepsilon_{Traditional}$$

- Transit use – binary logit

$$U_{Transit} = \alpha_0 + \alpha_1 \begin{pmatrix} Environmental \\ Consciousness \end{pmatrix} + \alpha_2 (Suburb) + \alpha_X \begin{pmatrix} Socio- \\ Demographics \end{pmatrix} + \varepsilon_{Transit}$$

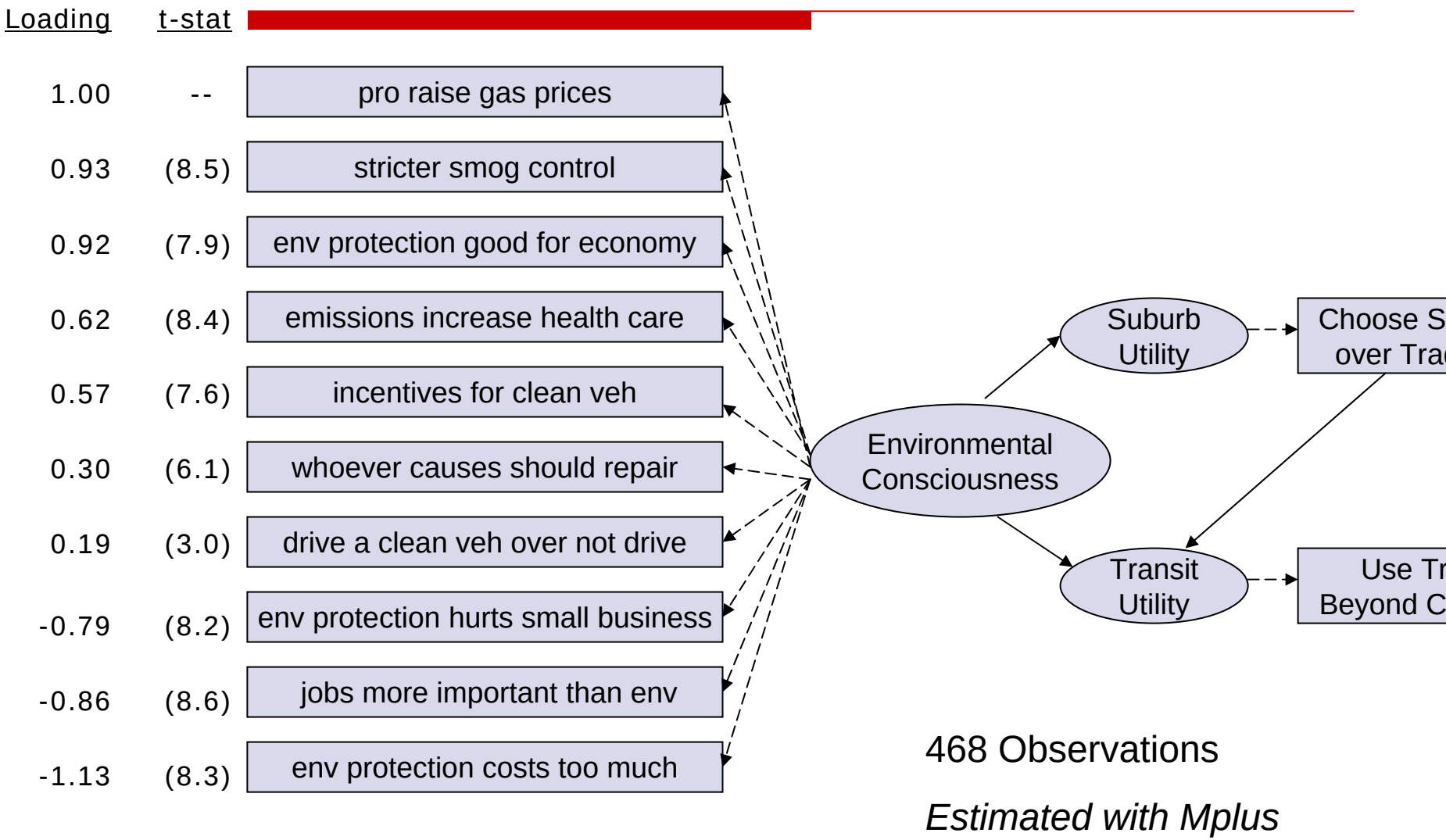
$$U_{NotTransit} = \varepsilon_{NotTransit}$$

- Environmentalism – set of linear eqs

$$Indicator_k = \lambda_k \begin{pmatrix} Environmental \\ Conciousness \end{pmatrix} + \varepsilon_k \qquad k = 1,...,10$$

$$\begin{pmatrix} Environmental \\ Conciousness \end{pmatrix} = \gamma \begin{pmatrix} Socio- \\ Demographics \end{pmatrix} + \varepsilon_{EC}$$

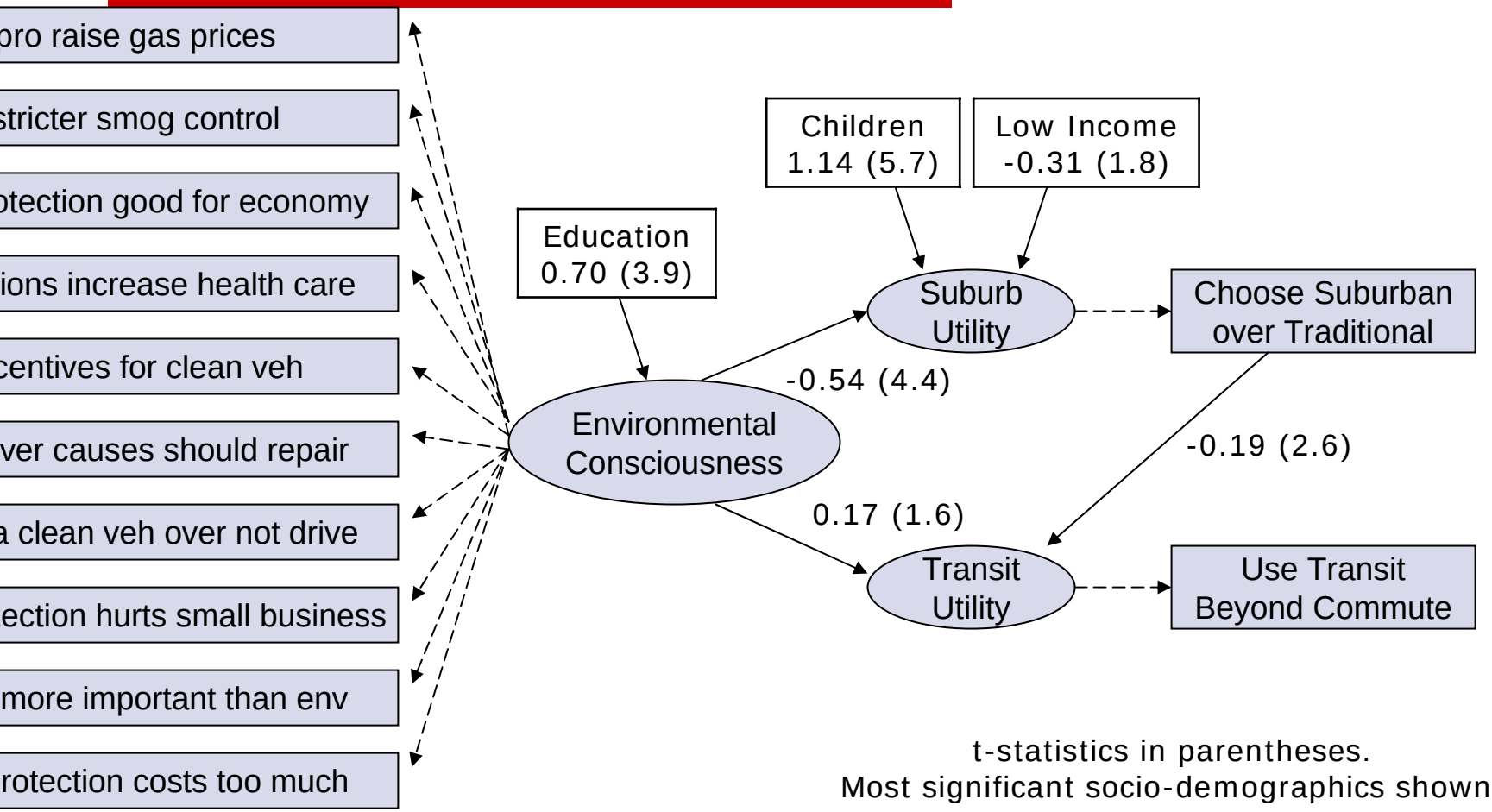
Estimation Results (1)



468 Observations

Estimated with Mplus

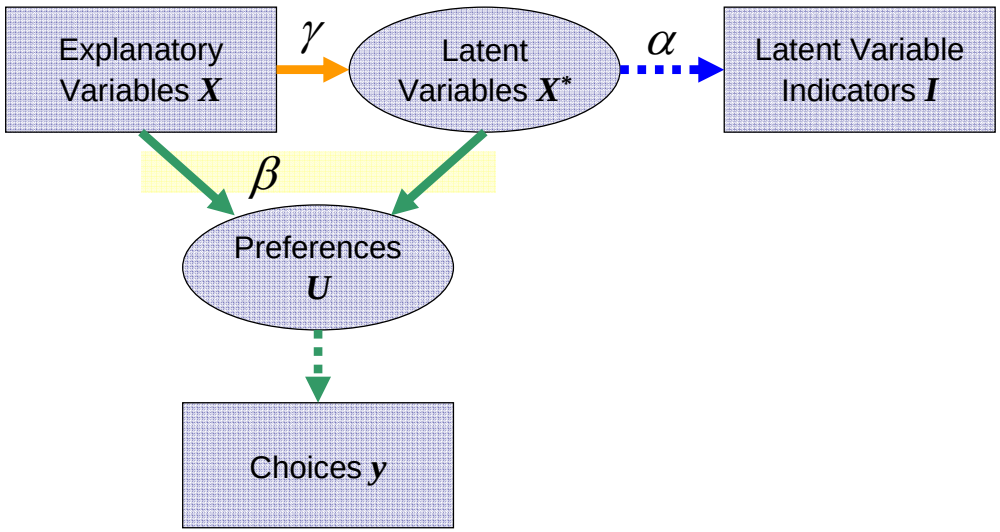
Estimation Results (2)



CONCLUSION

Method

- Integrated choice and latent variable models provide powerful & practical method to enrich behavioral representation in discrete choice models



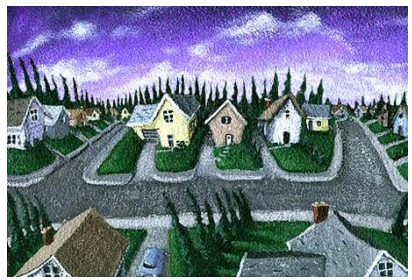
Comparison with State of the Art

1. Mixture Models
   - Sophisticated models of the covariance
     - Random parameters, error components
2. Behavioral Mixture Models
   - Model covariance structure via explicit latent variable constructs
   - Provide behavioral rational to mixtures

Fit?
Temporal stability over time?
Policy implications?

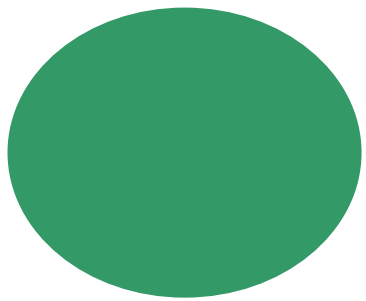
Applications Demonstrate More Intuitive Models

1. Lifestyle and Residential Choices



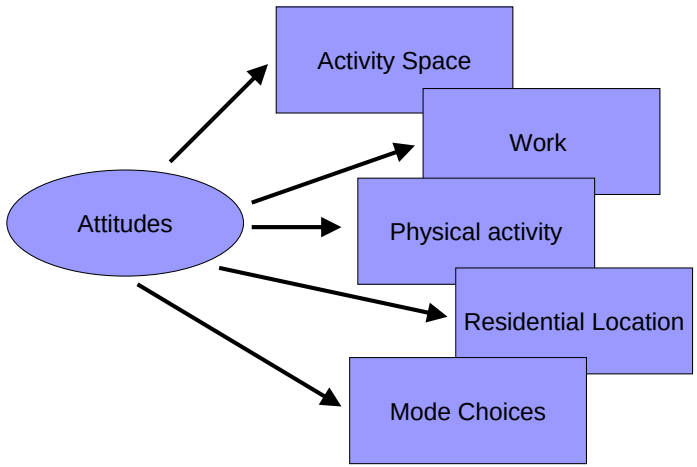
2. Risk, Rush and Airline Itinerary Choices

3. Role of Environmentalism



Future Directions

- Multi-contextual modeling



- Attitudinal trends
- Validation with forecasts