2-19-2016

Consistent Estimation of Route Choice Models for Dynamic Transit Assignment

Jeff Hood
Hood Transportation Consulting

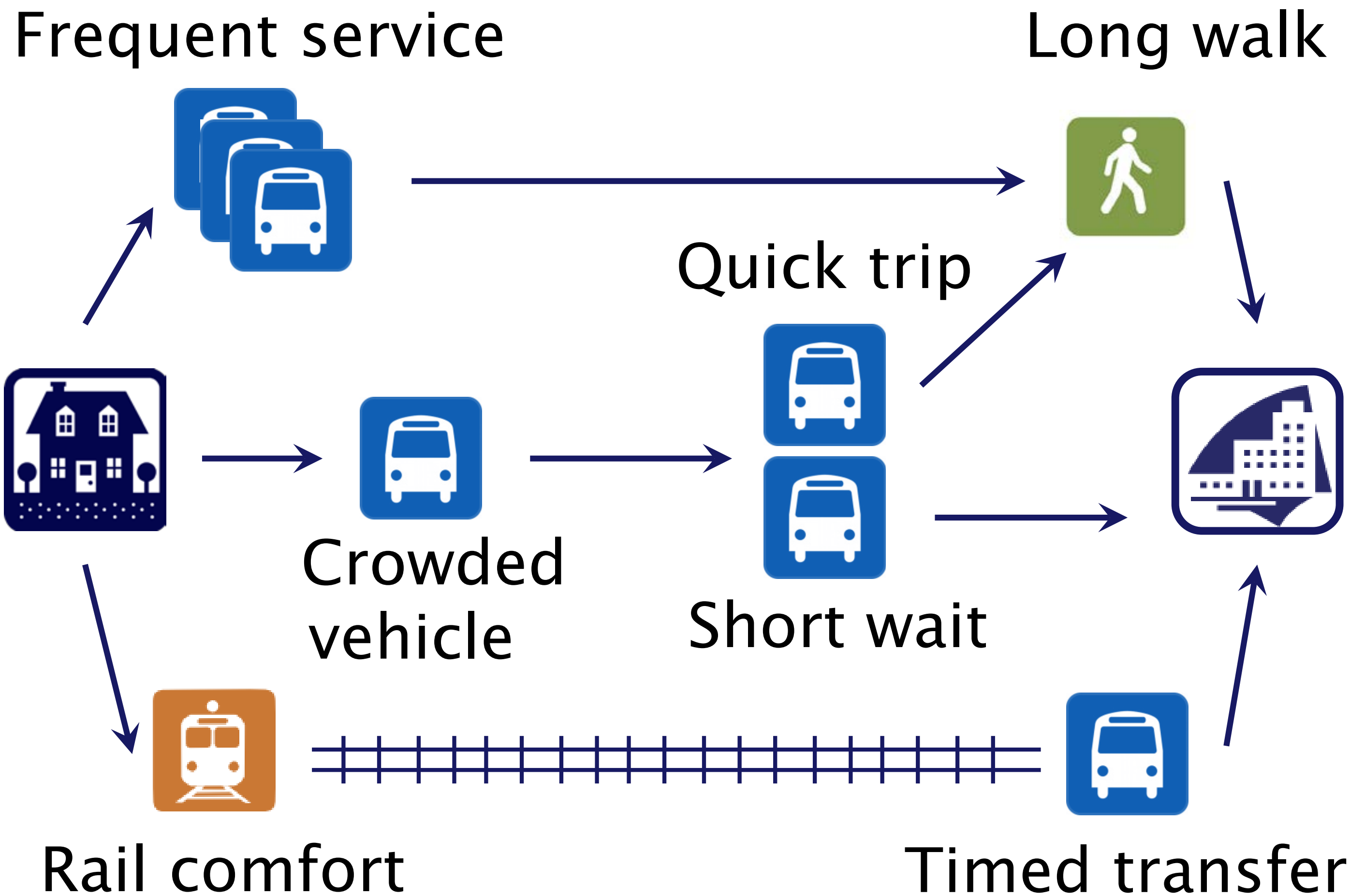
Consistent Estimation of Route Choice Models for Dynamic Transit Assignment

Jeff Hood

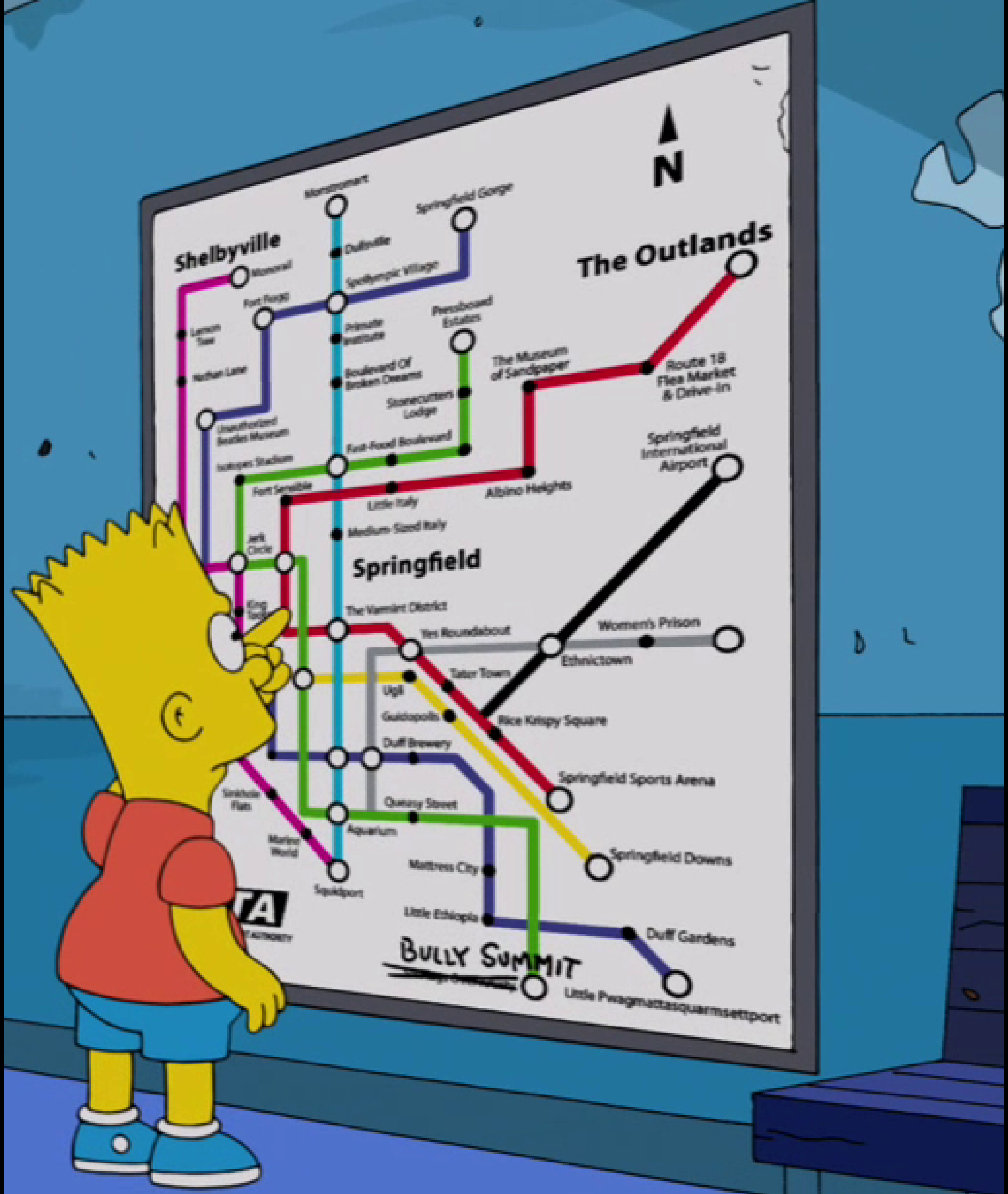
Hood Transportation Consulting

Portland State University Transportation Seminar
Portland, Oregon
February 16, 2016

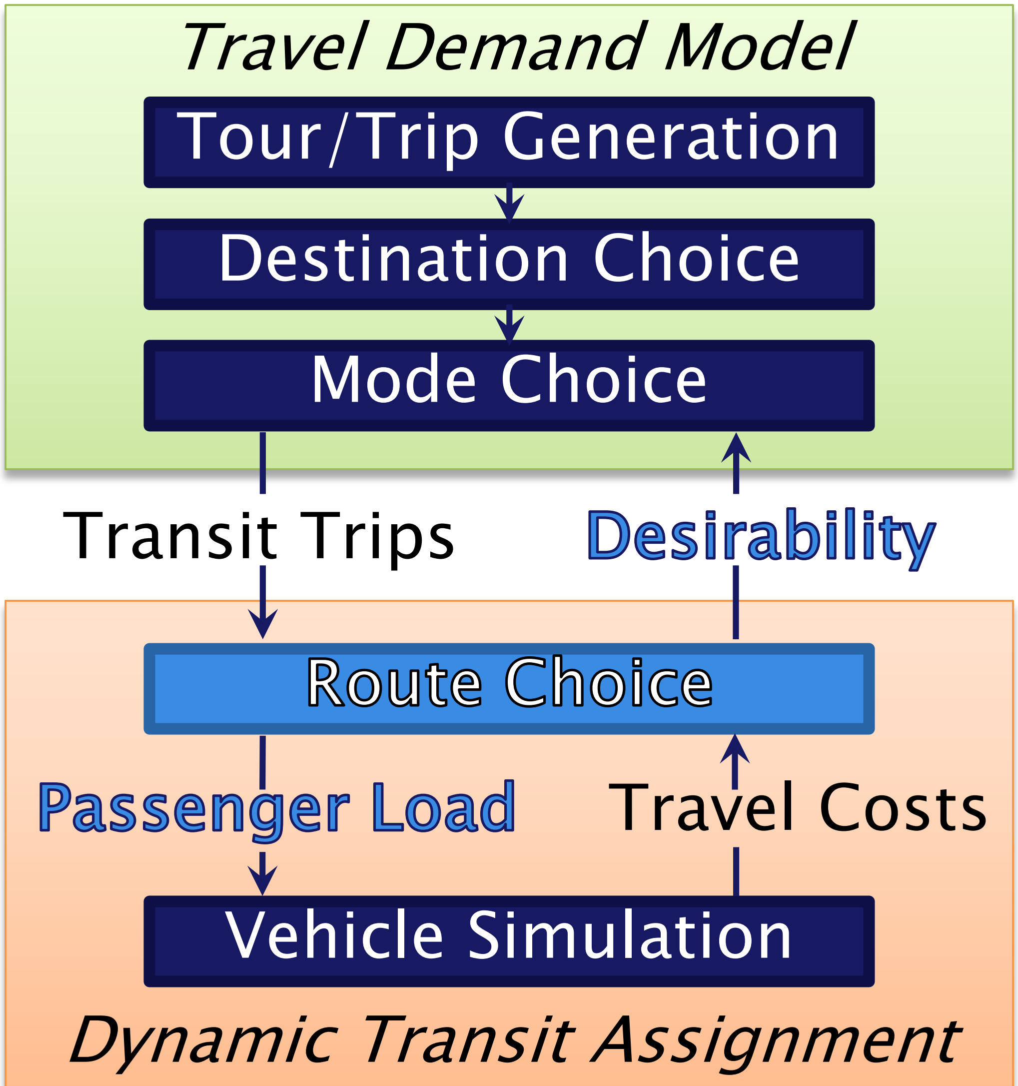
How do transit passengers choose a route?



Hood Transportation Consulting

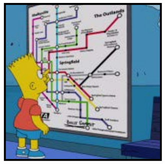

N
The Outlands
Shelbyville
Monument
Springfield Gorge
Dullsville
Spellympic Village
Monorail
Fort Fargo
Lemon Tree
Nuthan Lane
Private Institute
Pressboard Estates
Boulevard Of Broken Dreams
The Museum of Sandpaper
Route 18 Flea Market & Drive-In
Stonecutters Lodge
Unauthorized Beatles Museum
Fast-Food Boulevard
Isotopes Stadium
Springfield International Airport
Fort Sensible
Little Italy
Albino Heights
Jerk Circle
Medium-Sized Italy
Springfield
The Varmint District
Yes Roundabout
Women's Prison
Ethnictown
Tater Town
Ugli
Guidopolis
Rice Krispy Square
Duff Brewery
Springfield Sports Arena
Sinkhole Flats
Queasy Street
Aquarium
Marine World
Springfield Downs
Mattress City
Squidport
Little Ethiopia
Duff Gardens
BULLY SUMMIT
Little Pwagmattasquarmsettport

Predict passenger load and measure desirability in forecasting system

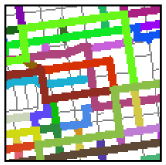


Hood Transportation Consulting

Outline



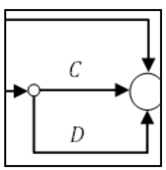
Introduction (you are here!)



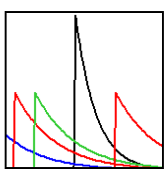
Limitations of path-based methods

$P(a|k)$

The recursive logit model

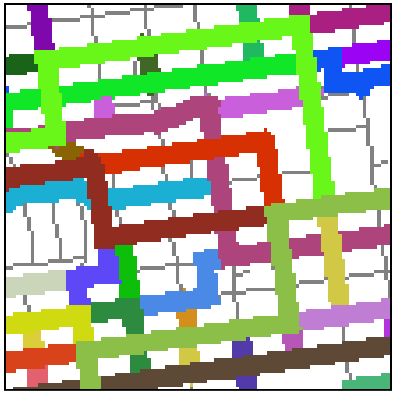


New correction for route overlap



Reliability and stochastic arrivals

Hood Transportation Consulting



Limitations of Path-Based Methods

Path-based models require choice set generation

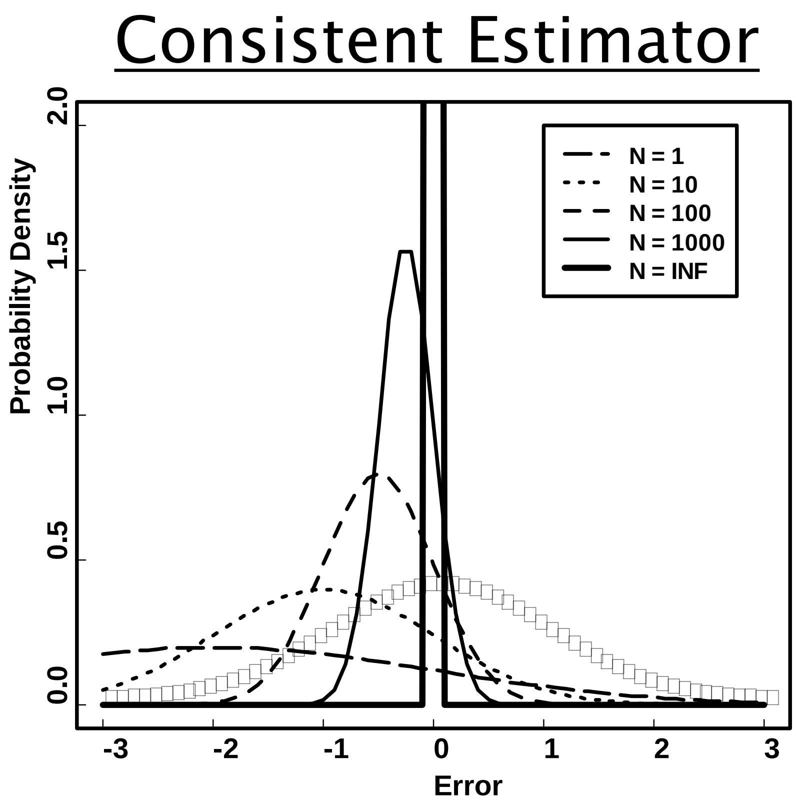
Hood Transportation Consulting

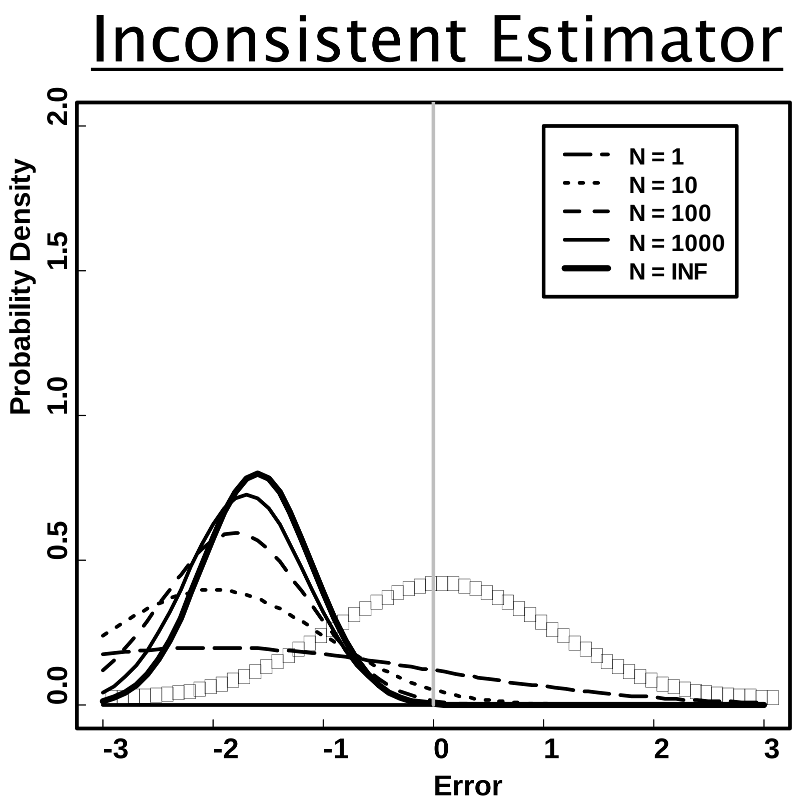
Example: stochastic sampling with unknown distribution

Hood et al. (2011). *Transport. Letters* 3,63–75.

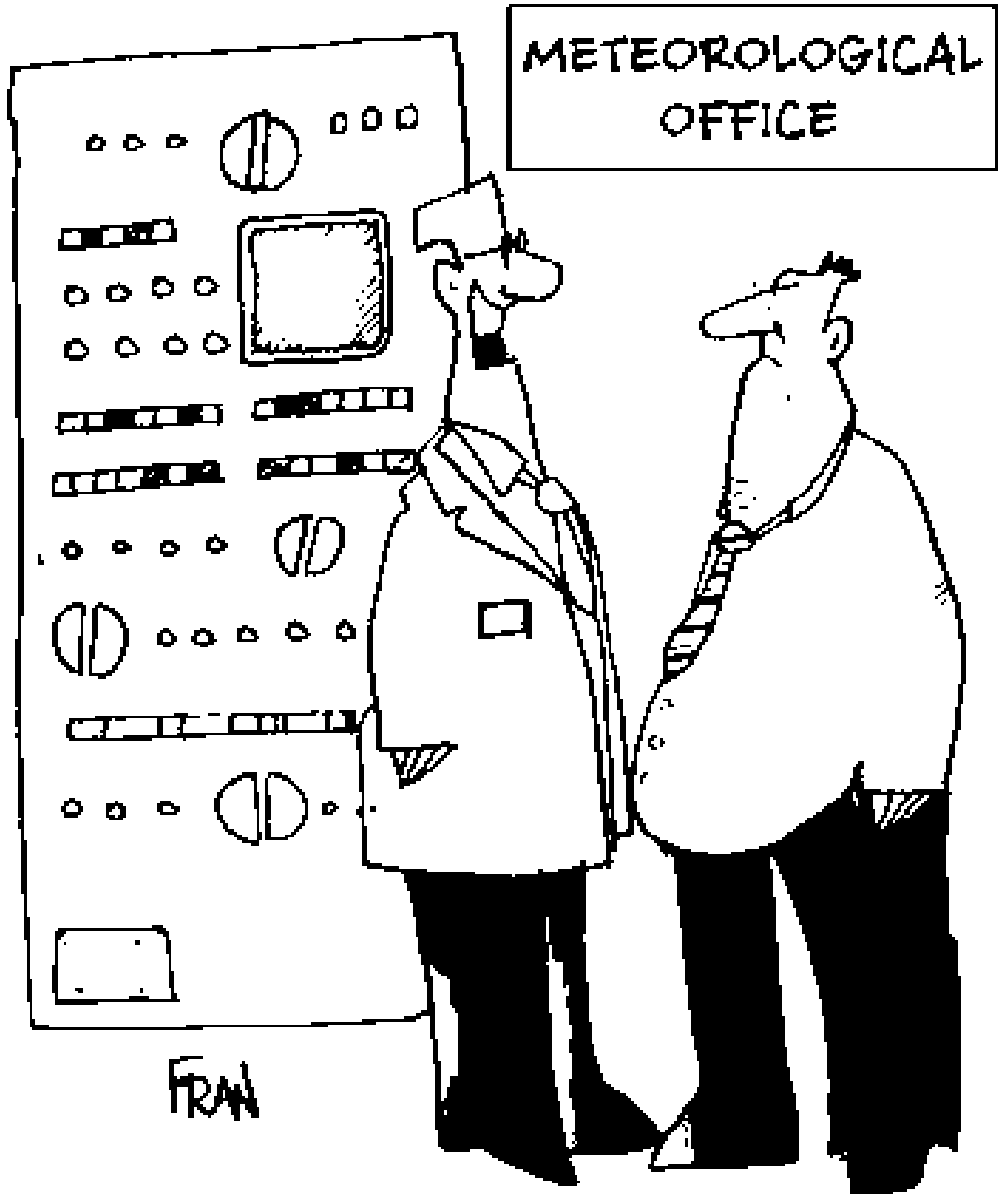
Hood Transportation Consulting

For most choice set generation schemes, estimators are inconsistent





Hood Transportation Consulting



With our new superfast computers we can get the forecast wrong TWICE as fast as we used to!

Consistent methods are too slow

Base Build Base Build

<u>Stochastic Sampling</u> <u>Path Size Link Penalty</u>

Nassir et al. (2014). *Transport. Res. Rec.* 2430, 170–181.

Run time is quadratic in zones.
1000 zones on 4 processors requires a week!

Hood Transportation Consulting

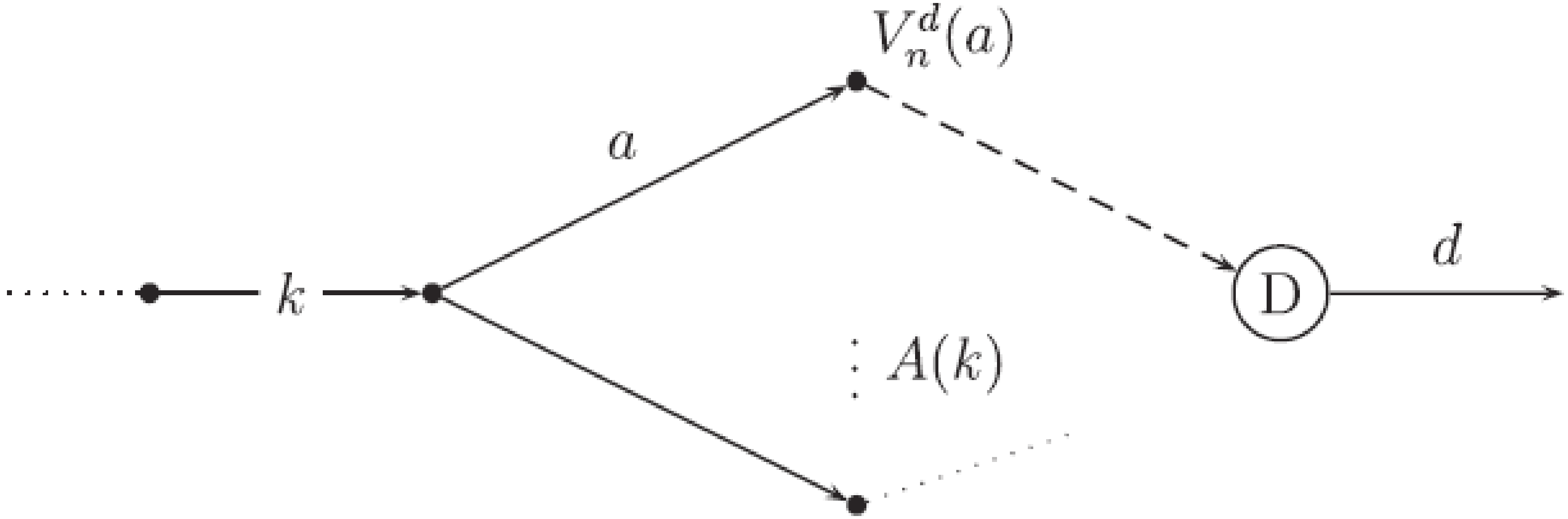

GOOD LUCK
45
M.P.H.


DEAD END
TURN BACK
YOU SHALL NOT PASS

$$P(a|k)$$

The Recursive Logit Model

Dynamic programming solution to Markov decision process



Fosgerau et al. (2014) *Transport. Res. B:* 56, 70–80

k: current link
a: possible movement from k
d: destination
$V(a)$: expected max. utility of all paths from a to d
$A(k)$: set of all successors of k
$v(a|k)$: “instantaneous” utility of moving from a to k

Traveler maximizes sum of instantaneous utility and expected utility to destination

Recursive value equation:

$$V(k) = E\left[\max_{a \in A(k)} \left(v(a|k) + V(a) + \mu\varepsilon(a)\right)\right]$$

<u>$\varepsilon(a)$ i.i.d. extreme value type I implies…</u>

Logit transition probabilities:

$$P(a|k) = \frac{\exp\frac{1}{\mu}[v(a|k)+V(a)]}{\sum_{a' \in A(k)} \exp\frac{1}{\mu}[v(a'|k)+V(a')]}$$

Value equation is logsum:

$$V(k) = \mu \log \sum_{a \in A(k)} \exp\frac{1}{\mu}[v(a|k) + V(a)]$$

Hood Transportation Consulting

Recursive equations solved efficiently with sparse linear system

If

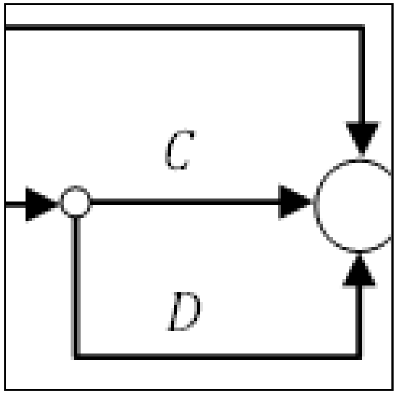
- $\mathbf{M}$ = matrix of exponentiated utilities $\exp[v(a|k)/\mu]$
- $\mathbf{b}$ = indicator vector for the destination
- $\mathbf{z}$ = desired vector of values $\exp[V(k)/\mu]$

Then the Bellman value equation is

$$\mathbf{z} = \mathbf{M}\mathbf{z} + \mathbf{b}$$

Result is equivalent to link-additive path-based model with unrestricted choice set

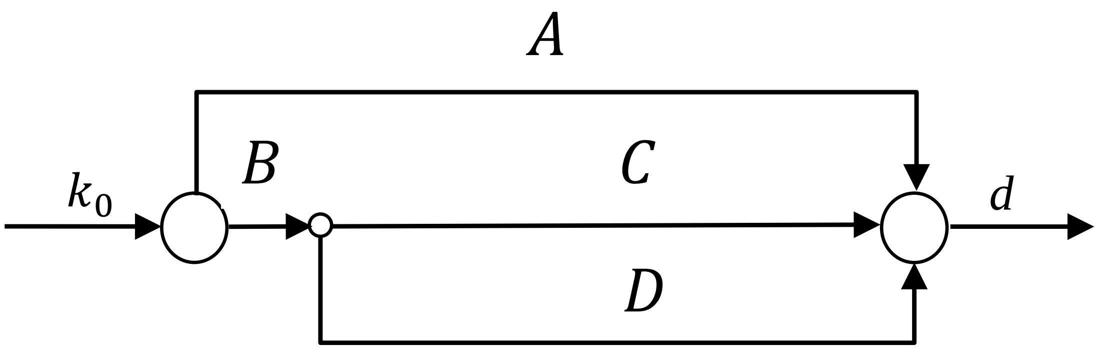
Hood Transportation Consulting



New Correction for Route Overlap

In reality, random errors are correlated due to overlapping routes

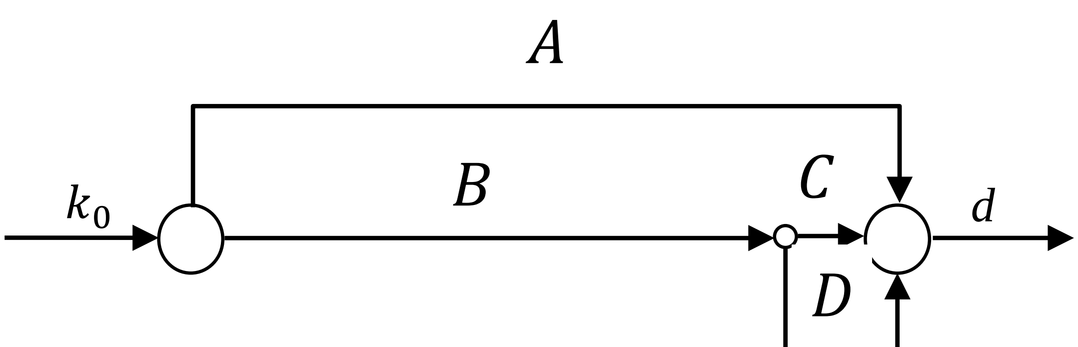
Uncorrelated Errors



Path-Based
"Size" Correction

Path	Size	Prob.
A	1	0.33
BC	1	0.33
BD	1	0.33

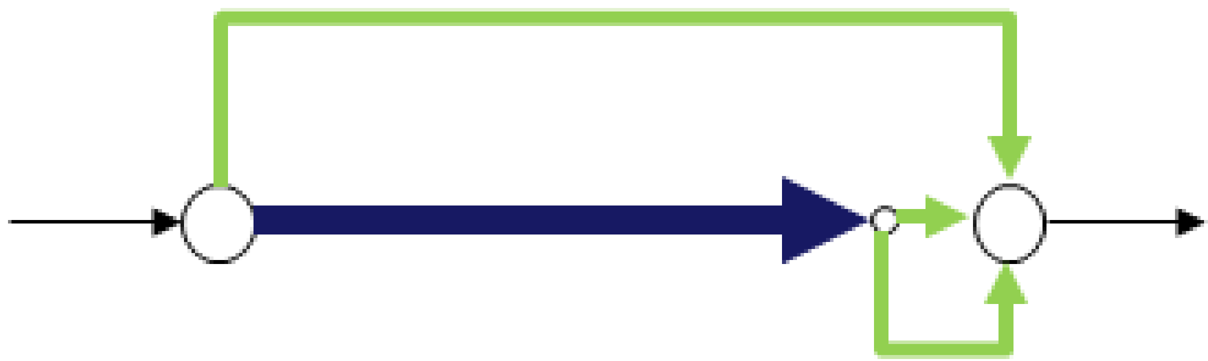
Correlated Errors



Path	Size	Prob.
A	1	0.50
BC	0.5	0.25
BD	0.5	0.25

Hood Transportation Consulting

For recursive logit, Fosgerau et al. recommend link flow proxy



<u>Performs Poorly</u>

- Not sensitive to extent of overlapping links
- Conflates route overlap and route utility
- Requires scaling parameter
- Not topologically-invariant

New link size variables extend recursive approach to counting of paths

downstream path segments

$$N^d(k) = \sum_{a \in A(k)} N^d(a)$$

upstream path segments

$$N^u(k) = \sum_{a \in A(k)} N^u(a)$$

paths containing k

$$N^d(k) \times N^u(k)$$

Equations have no solution in cyclic networks!

Hood Transportation Consulting

Path counts should be scaled by probability anyway

Probability-scaled downstream path segments

$$\tilde{N}^d(k) = \sum_{a \in A} \frac{P(a|k)}{\max_{a' \in A} P(a|k)} \tilde{N}^d(a)$$

Probability-scaled upstream path segments

$$\tilde{N}^u(k) = \sum_{a \in A} \frac{F(a) P(k|a)}{\max_{a' \in A} F(a') P(a')} \tilde{N}^u(a)$$

($F(a)$ is uncorrected link flow)

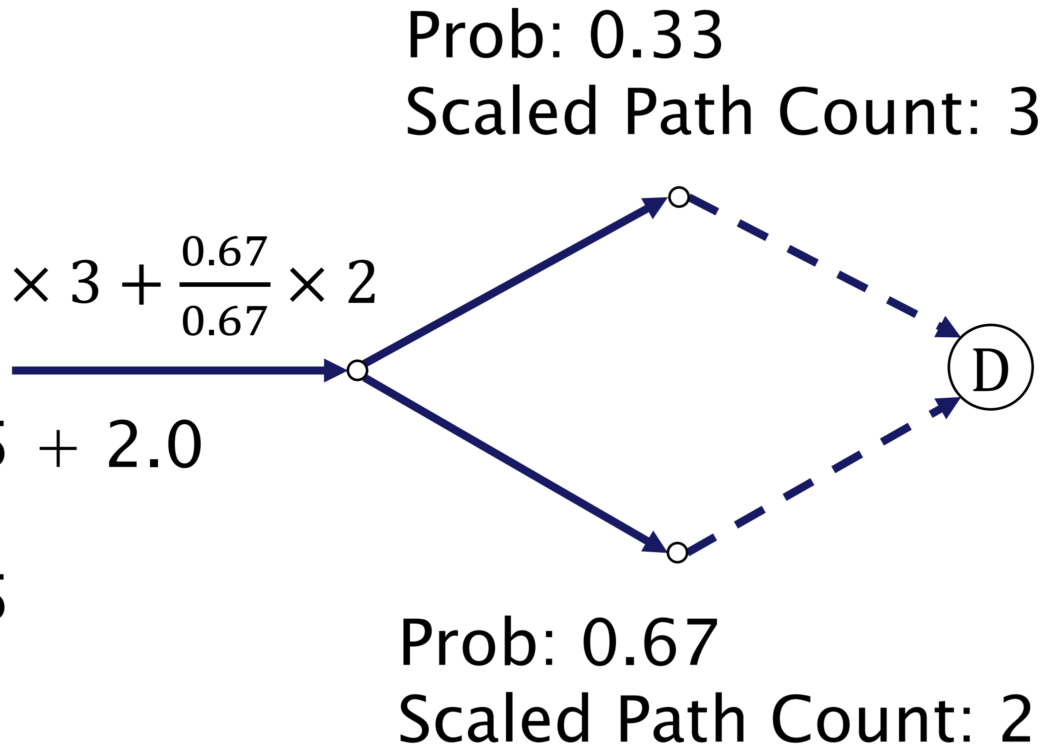
Hood Transportation Consulting

Example: scaled path count recursion

$$\text{S.P.C.} = \frac{0.33}{0.67} \times 3 + \frac{0.67}{0.67} \times 2$$

$$= 1.5 + 2.0$$

$$= 3.5$$



Hood Transportation Consulting

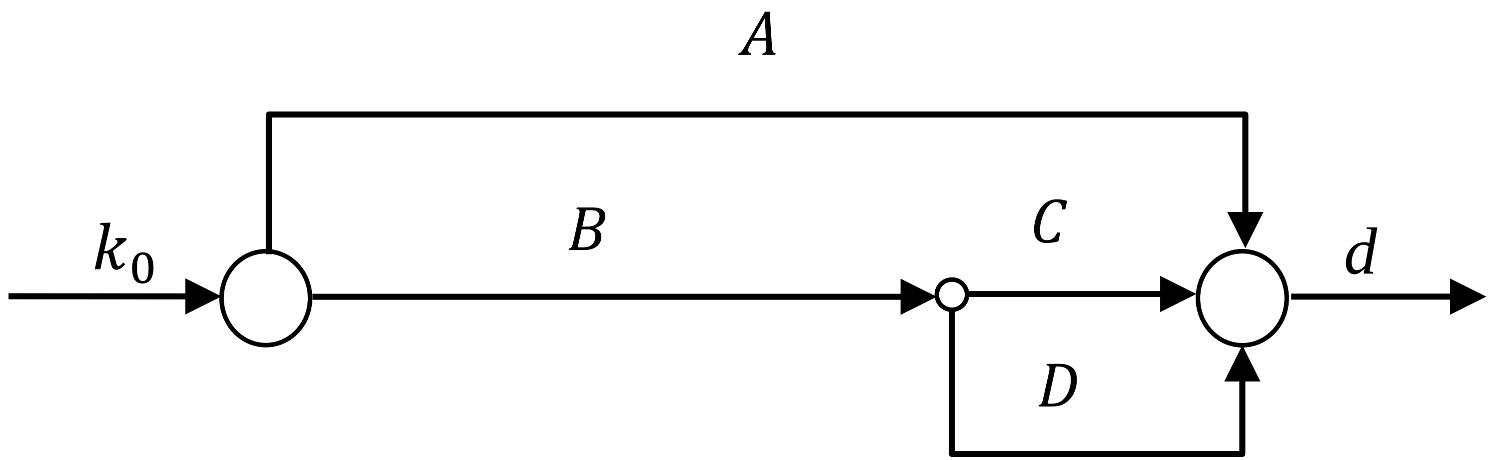
Link size variable follows established form from path-based methods

$$LS(k) = \frac{m(k)}{\widehat{M}(k)} \log \widetilde{N}^{d}(k) \widetilde{N}^{u}(k)$$

$m(k)$: measure of link extent
$\widehat{M}(k)$: expected measure of all paths containing k

Hood Transportation Consulting

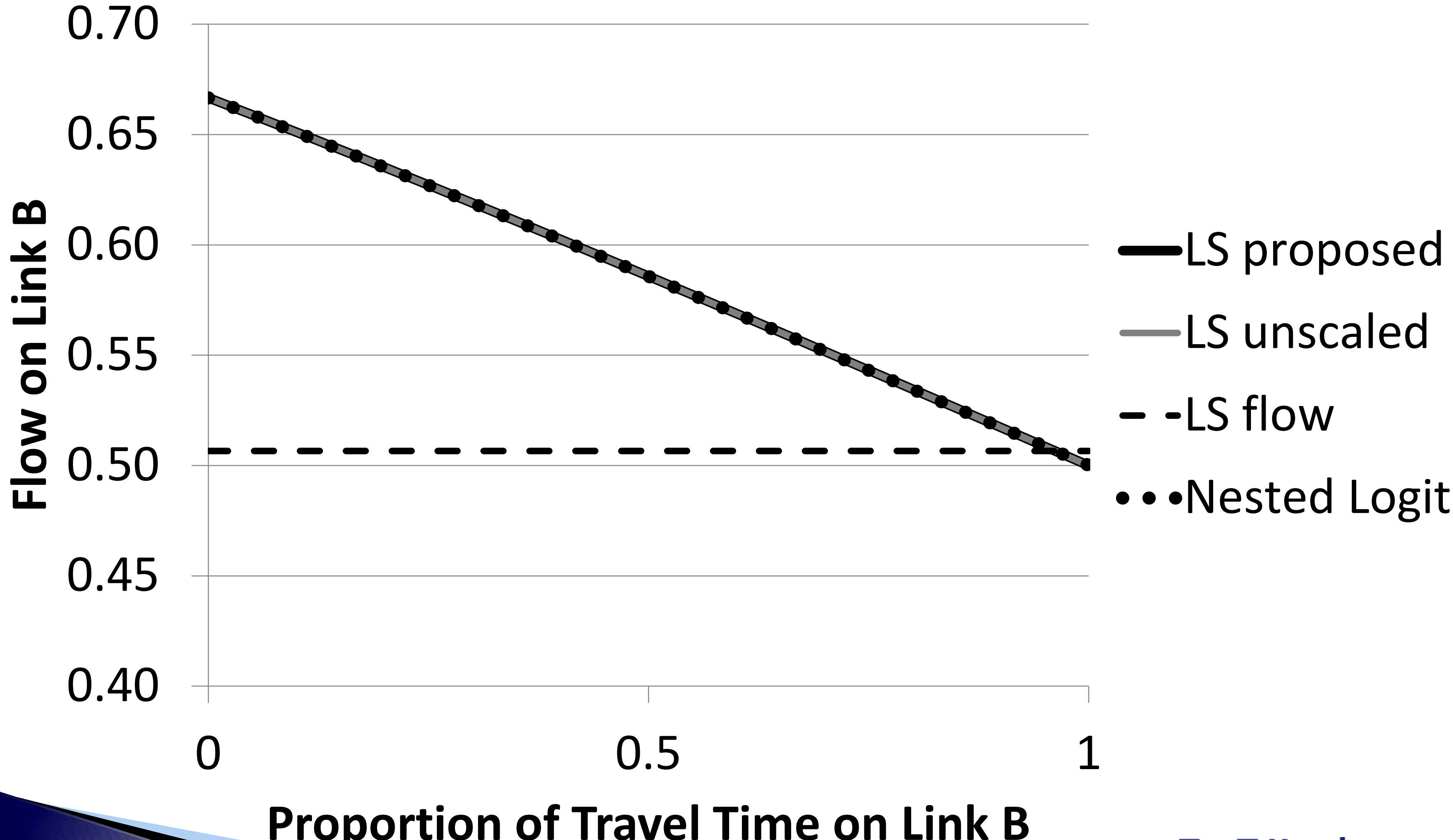
Example: collapsing downstream alternatives



Link	Measure $m(a)$	Travel Time
A	1	1
B	t	t
C	$1-t$	$1-t$
D	$1-t$	$1-t$

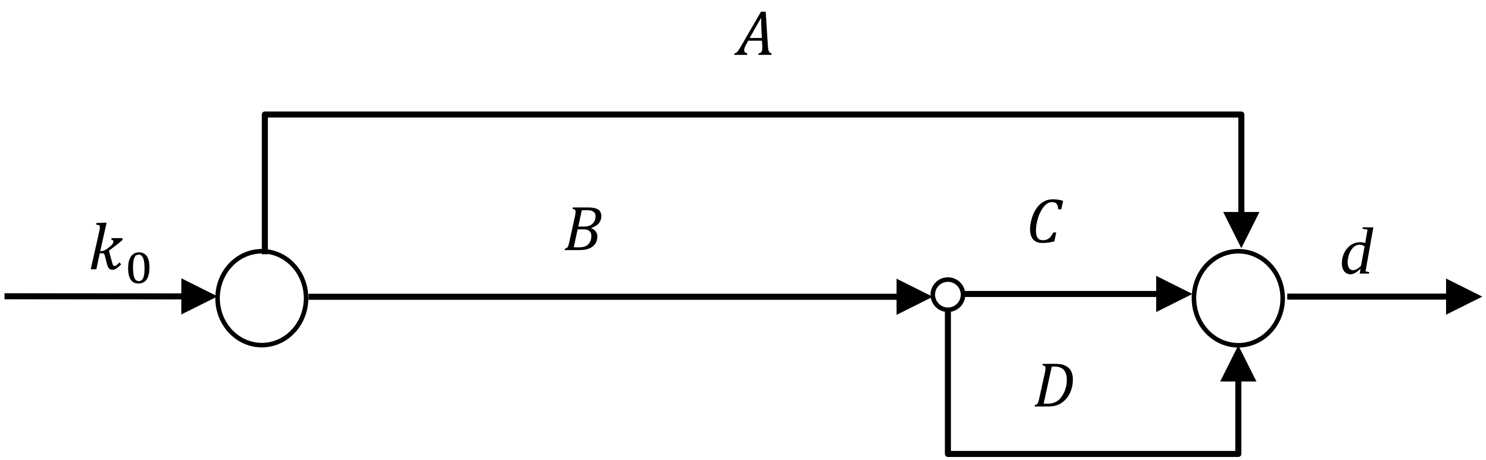
Hood Transportation Consulting

Example: collapsing downstream alternatives



Hood Transportation Consulting

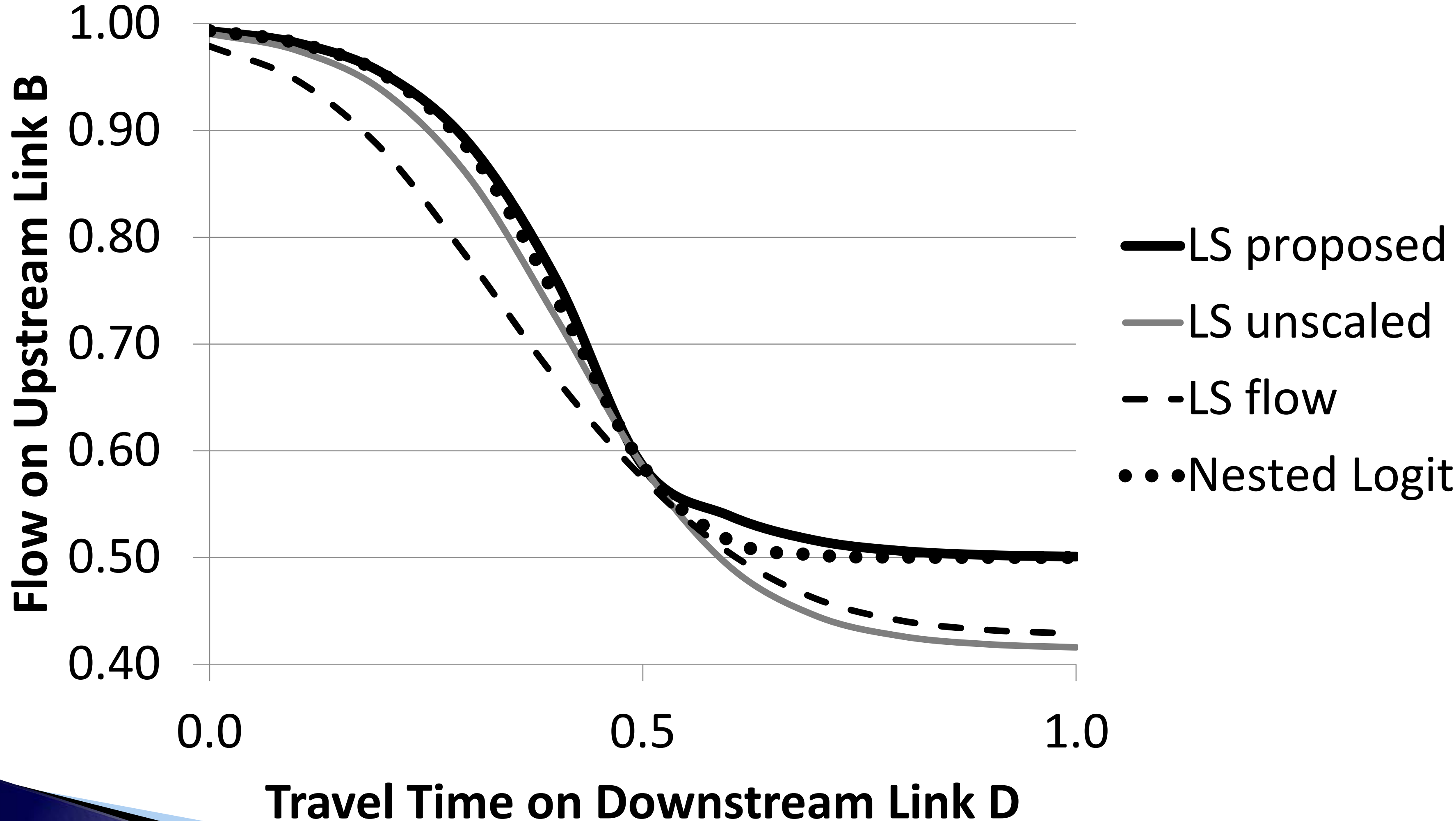
Example: varying downstream utilities



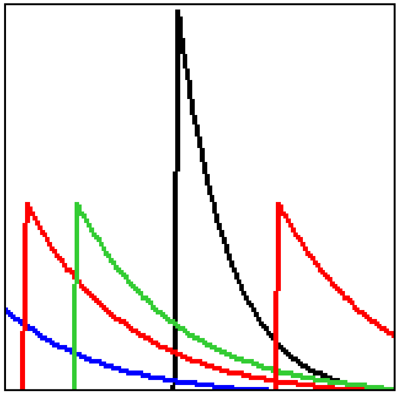
Link	Measure $m(a)$	Travel Time
A	1	1
B	0.5	0.5
C	0.5	0.5
D	t	t

Hood Transportation Consulting

Example: varying downstream utilities



Hood Transportation Consulting



Reliability and Stochastic Arrivals

Hood Transportation Consulting

Traveler responses to reliability are more complex for transit

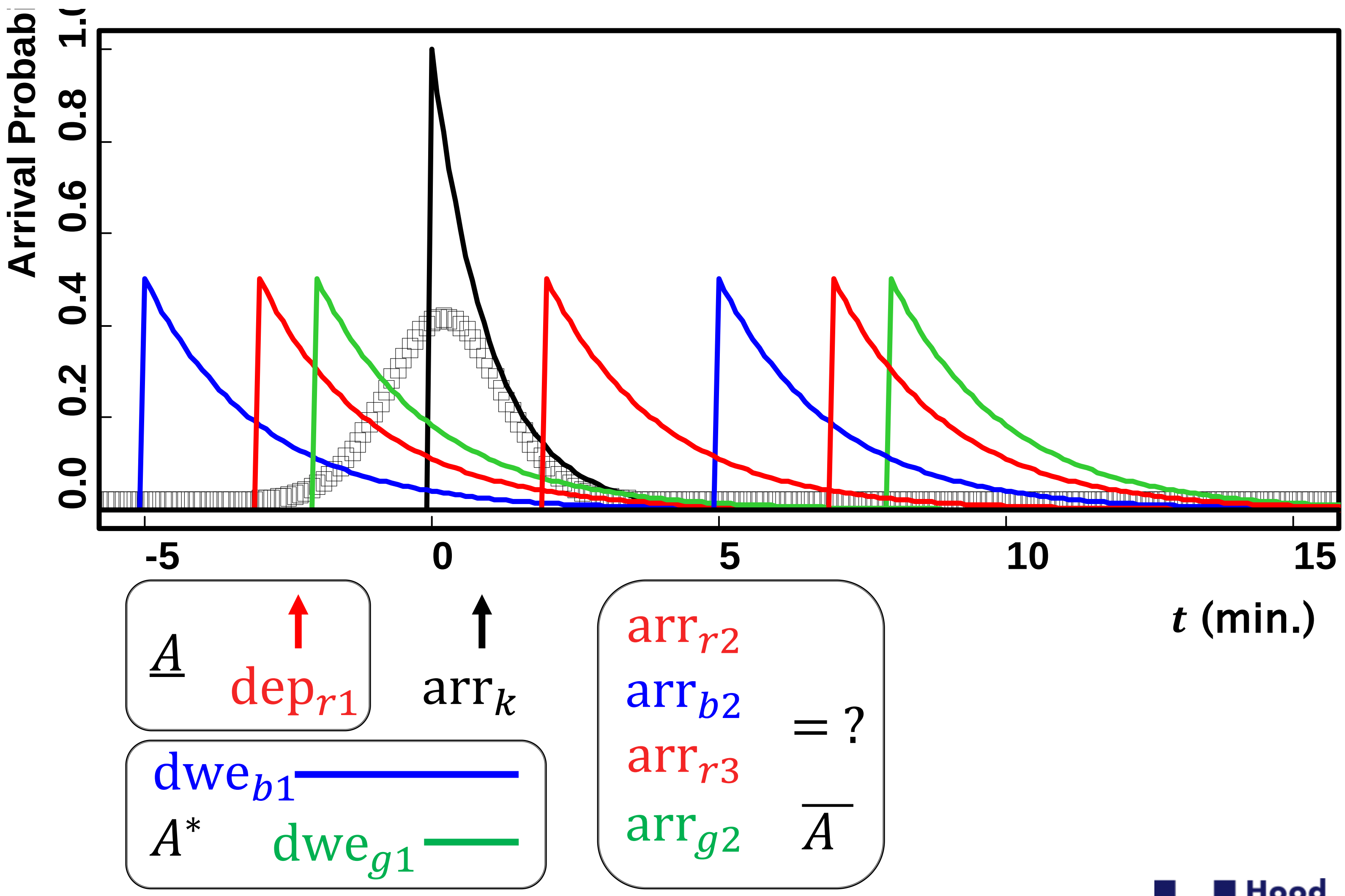
Poor schedule adherence reduces boarding probability for multiple reasons

- Direct disutility of excess wait times
- Missed connections
- Lateness in arrival sequence

"Reliability" term in utility function will not work

Problem requires **dynamic choice probabilities**

Boarding is conditional on which vehicles have departed, are dwelling, or yet-to-arrive



Hood Transportation Consulting

Probability of boarding a dwelling vehicle is

$$P\left(a \middle| k, \bar{A}, a \in A^*\right) = \frac{e^{\frac{1}{\mu}[\beta_{\text{dwe}} E(\text{dwe}_a) + v(a|k) + V(a)]}}{\sum_{a' \in A^*(k)} e^{\frac{1}{\mu}[\beta_{\text{dwe}} E(\text{dwe}_{a'}) + v(a'|k) + V(a')]} + e^{\frac{1}{\mu} E(U_{\text{wait}}|k, \bar{A})}}$$

Depends on expected utility of waiting $E\left(U_{\text{wait}} \middle| k, \bar{A}\right)$

Hood Transportation Consulting

What is the expected utility of waiting?

<u>Recursive formula depending on</u>

- Conditional distribution of arr_k given $\bar{A}$

$$\Phi_k(t|\bar{A}) = P(\mathrm{arr}_k < t|\bar{A})$$

- Conditional probability that next arrival is a_i

$$P\left(\min_{a_j \in \bar{A}} \mathrm{arr}_{a_j} = \mathrm{arr}_{a_i}\right)$$

- Expected utility of waiting for a_i

$$E(w(a_i|k, \mathrm{arr}_k)) = \int_0^\infty w(a_i|k, t^+) d\widetilde{\Phi}_{a^+}(t|A^+)$$

Hood Transportation Consulting

Probability of boarding a later vehicle given the decision to wait is

$$P(a|\text{wait}(\bar{A}))$$

$$= \sum_{a_i \in \bar{A}} P\left(\min_{a_j \in \bar{A}} \text{arr}_{a_j} = \text{arr}_{a_i}\right)$$

$$\times \begin{pmatrix} \delta(a = a_i) P\big(\text{board}(\bar{A}\backslash\{a_i\})\big) \\ + \delta(a \neq a_i) P\big(\text{wait}(\bar{A}\backslash\{a_i\})\big) P(a|\text{wait}(\bar{A}\backslash\{a_i\})) \end{pmatrix}$$

Hood Transportation Consulting

Marginal probability of boarding vehicle a is

$$P(a|k) = \sum_{i=0}^{|A(k)|} \sum_{j=0}^{i} \sum_{\underline{A} \in C(A(k),i)} \sum_{\bar{A} \in C\left(A(k) \setminus \underline{A}, j\right)} P\left(\underline{A}, \bar{A}, A^*\right)$$

$$\times \left(\begin{array}{c} \delta(a \in A^*) \\ \times P\left(a \middle| k, \bar{A}, a \in A^*\right) \\ + \delta(a \in \bar{A}) \left(1 - \sum_{a' \in A^*} P\left(a \middle| k, \bar{A}, a \in A^*\right) \right) \\ \times P(a|\text{wait } (\bar{A})) \end{array} \right)$$

If delays are independent exponentials, there is an (excruciating) closed-form solution

Hood Transportation Consulting

Kathryn Coffel
David Ory
Lisa Zorn
Shimon Israel

Suzanne Childress
Stefan Coe
Joe Castiglione
Drew Cooper

Elizabeth Sall
Alireza Khani
Emma Frejinger
Tien Mai

Thank you!

Contact: jeff@hoodconsulting.net

Hood Transportation Consulting