Crossroads of travel modeling today

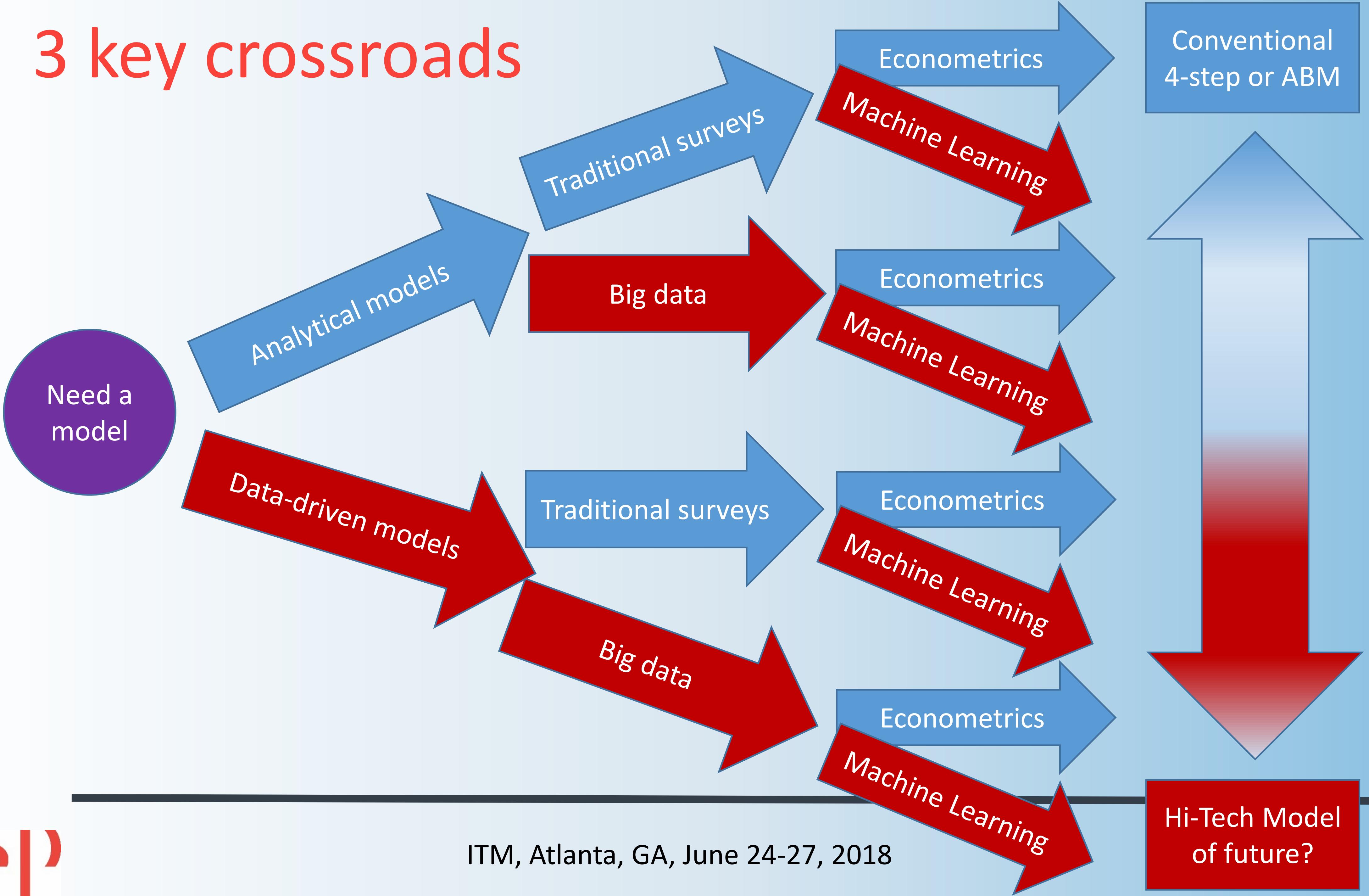
Peter Vovsha (WSP)

Session 4C

"The Battle of the Methods – Machine/Deep Learning vs Econometric/Statistical Modeling"

ITM, Atlanta, GA, June 24-27, 2018

3 key crossroads



Data-driven vs. analytical

Use model to predict growth or incremental change pivoting off the data

Basic differences

Analytical

- Use trip patterns to estimate model coefficients
- Rely on analytics to generate trip patterns from scratch
- Applied identically to base year and forecasting

Data-driven

- Use trip patterns directly to the maximum extent
- Minimize data transformation to smoothing and expansion
- Need growth factors and re-expansion for future years

Other angles of view

Data-driven models

- Essentially non-parametric over-specified structures w/zillions of constants
- Still need some analytics for forecasting:
  - Re-expansion (growth factors) for future years
  - At minimum mode choice and/or route-choice needed for policy sensitivity

Analytical models

- Somewhat data-driven through model estimation (training)
- Somewhat over-specified in the model calibration:
  - K-factors in trip distribution
  - Mode-specific constants by geography
  - Trip matrix adjustment to traffic counts

Intuitive applications of data-driven concept in the past

- Trip matrix balancing:
  - Spatial trip distribution (destination choice) is the most difficult travel dimension to tackle analytically
- Incremental logit models for mode choice:
  - Baseline mode choice can be well established by adjustment of auto trips table to traffic counts and transit trip table from OB survey
  - Incremental mode choice would always generate a logical modal shift
- Typical traffic engineering approach:
  - Take initial vehicle trip table from regional model or “big data”
  - Adjust to traffic counts
  - Randomize trip departure time
  - Rely on traffic simulation or DTA (w/route choice) to evaluate projects and polices

Hybrid balancing-gravity: data-driven & analytical

OD-Survey

$\downarrow$

Impedance c_{ij} | Seed Matrix s_{ij}

$\downarrow$ | $\downarrow$

Gravity Model | Proportional Balancing

$$x_{ij} = P_i \alpha_i A_j \beta_j \psi(c_{ij})$$

$$x_{ij} = P_i \alpha_i A_j \beta_j s_{ij}$$

Combined Model

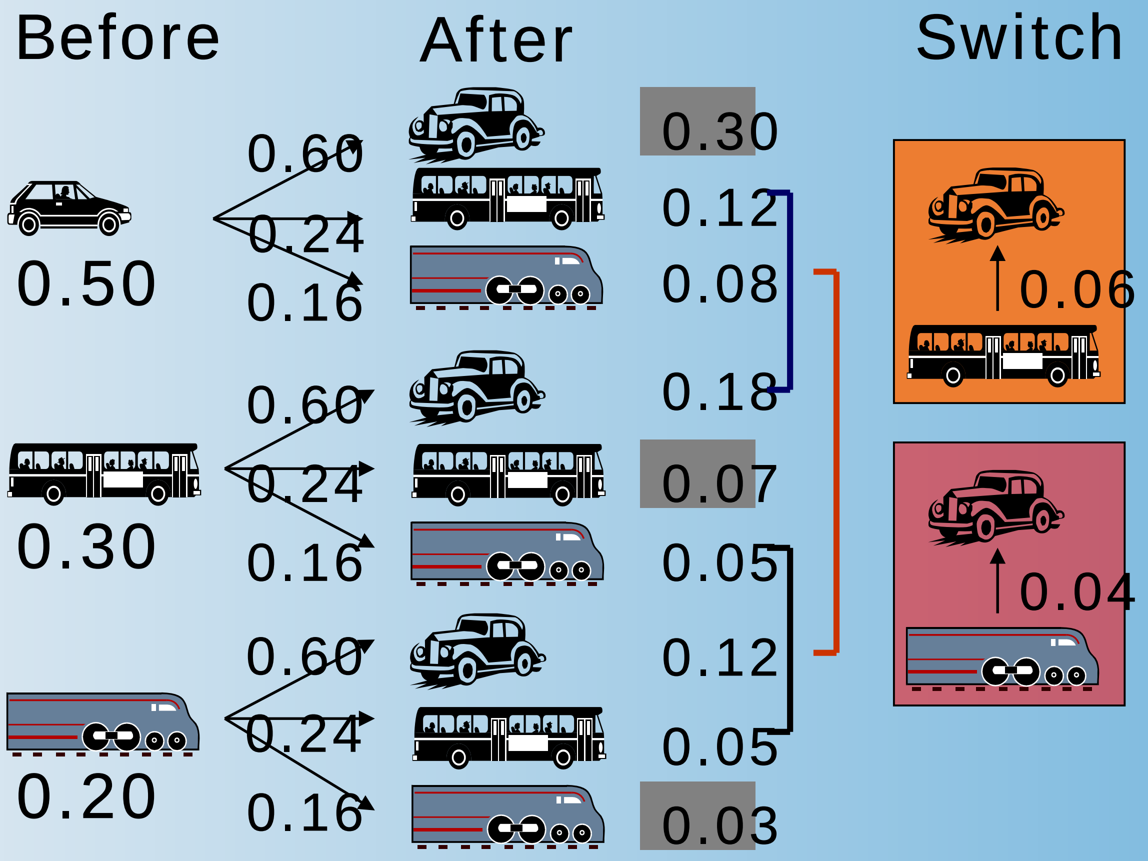
$$x_{ij} = P_i \alpha_i A_j \beta_j s_{ij}^{(\mu_i+\nu_j)/2} \psi(c_{ij})^{1-(\mu_i+\nu_j)/2}$$

Incrementality & pivoting

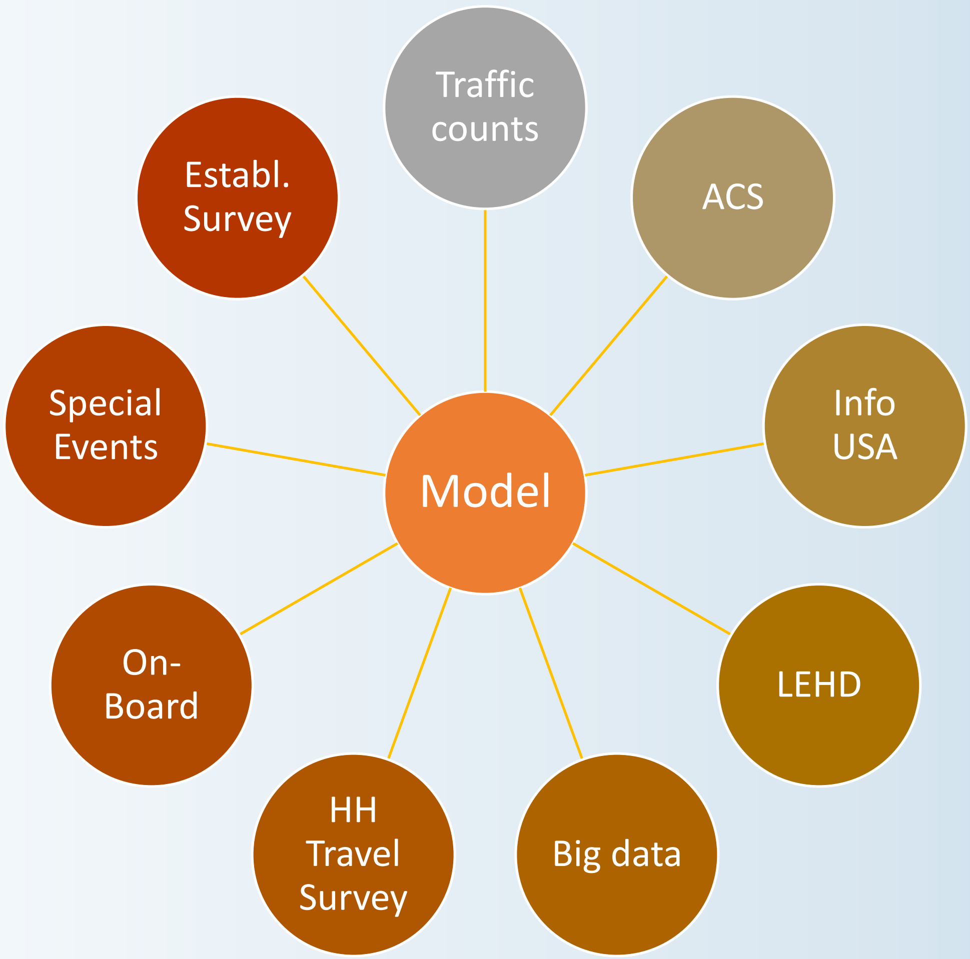
- Primarily for spatial distribution
- Trip generation as well
- Incremental logit mode choice and TOD choice:
  - Switching logit model for microsimulation
- Can it be extended to DAP?
  - MatSim concept (genetic algorithm)
  - C10 concept of gradual “freezing” of HHs (stressed/unstressed)

Switching Logit Model (more details in Session 6B)

- Generalization of Incremental Logit:
  - No base case calibration
  - Standard Incremental Logit does not work with individual records
  - Switching Logit is a theoretically sound construct that does the trick
- Explicitly model mode switch:
  - Previous (observed) mode is known
  - Switching probabilities are consistent with the estimated core model
- Clarification:
  - Switching Logit is the way of model application
  - W/o transaction cost it is estimated as ordinary Logit



New view on travel model as data integration tool



- MPOs/DOTs collects many different types of data
- Travel model is the lowest possible denominator
- Travel model incorporates all surveys and consolidates them in one consistent output

Big Data vs. conventional surveys

How can we take a full advantage of passively collected Big Data beyond just travel model validation

Big Data invasion

- Big data has many advantages but can it be used beyond model validation?
- Big data is not behavioral:
  - Detailed trip purpose?
  - Other modes?
  - Household income?
  - Person type and age?
  - Trip chaining and linkage to person?

Behavior-ization of Big Data – closing the gap

- Trip purpose:
  - Can be imputed based on detailed parcel-level land use and temporal profile of activity over multiple days
  - Travel model can be reformulated in terms of activity profiles and establishment types rather than traditional trip purpose labels:
    - Instead of “shopping” trip purpose model could operate with a “recurrent activity of 2 hours or less at a shopping center with a weekly cycle”
- Trip mode:
  - Can be reliably imputed based on speed profile and comparison of itinerary to the detailed road and transit networks
- Household income:
  - Can be replaced with the average residential zone income as an explanatory variable that in several ABMs performed statistically better than individual household income.
- Person type and age:
  - Deduced from activity types and levels over multiple days
  - Travel model can be reformulated in terms of activity levels rather than person types and age categories

Retaining person ID over multiple trips and days is the key

- Sidewalk Labs "Replica"
- Observed individual daily patterns can be classified and linked to synthetic population
- Can be a breakthrough in combination with incremental/pivoting/switching ABM

ML vs. Econometrics

ML is teaching a machine to do what people can do easily

2 lines of research

- ML as replacement for logit models within the same model system design:
  - Classifiers instead of discrete choice
  - Penalized likelihood with priors and learning
  - Trials underway
- ML as principal revolution in the model structure:
  - Neural networks?
  - Pattern recognition for relating big data to synthetic population
  - Need more time to evaluate potential and directions

“Secrets” of ML match

Model	Major distinct feature
Neural network	Box-cox type logistic transformation of groups of variables and mapping to [0,1]
Random forest	Thresholds of variables and combinations of them
SVM	Kernels (combinations of variables) and separation of common cases from outliers
K-neighbor classifiers	Non-linear combinations of variables by regions of their values
Common to all ML	Model specification and estimation are integrated in one automated process

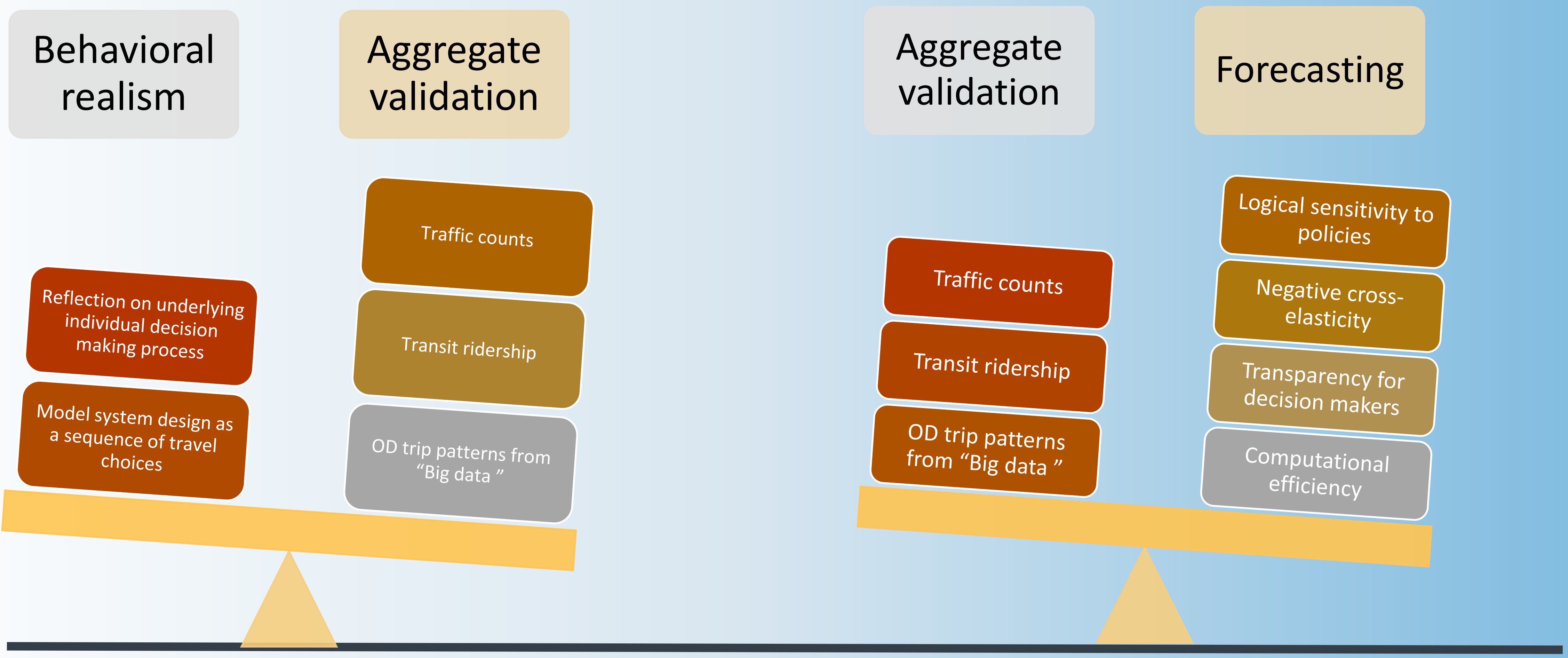
Important concerns to address with ML

- Expected logical properties of choice models:
  - Positive direct elasticity
  - Negative cross-elasticity
  - Sensitivity to policies in expected and controlled range
- All this is not automatically held for ML methods:
  - New toll road? And cost was not significant?
- Traditional high-tech-driven ML approaches are based on a very different notion of forecasting or predicting or dynamics:
  - ML predicting is similar to our Jack-knife validation
  - Dynamics are mostly real-time and rarely evolutionary (except for some AgBMs)
  - Sensitivity to policies (or changing environment) is not in the focus of ML

Conclusions

- Evolution or revolution?
  - Many useful hybrids emerge depending on the project
- Fine line between conservatism and jumping on shiny innovations:
  - Only detailed analysis and application experience can tell
- Forecasting is not exactly hi-tech:
  - Too open w/elements of art in addition to science
  - Inputs and outputs are not always well-defined

Evaluation of travel models: different aspects of “truth”



Contact(s)

Peter Vovsha, PhD

Assistant Vice President, WSP

Systems Analysis Group

Peter.Vovsha@wsp.com

Formerly

WSP | PARSONS BRINCKERHOFF

Question the existing

Imagine the impossible

Create the enduring

wsp.com/usa