Disaggregate simulation and some implications for calibration

Gunnar Flötteröd

June 28, 2011
An appeal for a more disaggregate perspective

• common sense is to “keep things simple”
• increased disaggregation
  – looks more at the “distributed” side of a simulation
  – may appear as if one was not keeping things simple
• this talk indicates relevance and feasibility of disaggregation
One argument: calibration principles stay clear

• calibration problem statement:

\[ P(X \mid Y) \propto P(Y \mid X)P(X) \]

• naive simulation of the solution:

\[
E\{X \mid Y\} \propto \int XP(Y \mid X)P(X)dX \\
\approx \frac{1}{R} \sum_{r=1}^{R} X' P(Y \mid X'); \quad X' \sim P(X)
\]

• It is possible to do things like this for very large systems!
Outline

Relevance: truthful modeling of uncertainty

Doability: avoiding to drown in details

Summary
Outline

Relevance: truthful modeling of uncertainty

Doability: avoiding to drown in details

Summary
Explaining deviations from reality

- when calibrating a (complicated) microsimulation...
  - one needs some kind of **calibration model**
  - this model must explain all deviations from reality

- essentially two approaches:
  1. use a deterministic calibration model\(^1\) and add random slack
  2. explicitly use a stochastic model to represent uncertainty

\(^1\) assignment matrix, OD matrix, linear dynamics, response surface, ...
Deterministic calibration model + random slack?

• (typical) measurement equation:

\[ Y = F(X_1) + \varepsilon(X_2) \]

• two extreme cases
  
  1. analyst really knows what is going on: \( Y = F(X_1) \)
  2. analyst does not get the causality right: \( Y = \varepsilon(X_2) \)

• in transportation, one seems to deal more with case 2...
Example: choice set uncertainties

• (simulated) travel behavior with uncertain choice sets:

\[ P_n(i) = \sum_{C_n \in \mathcal{C}} P_n(i \mid C_n) P_n(C_n) \]

• operational version:

\[ P_n(i) = P_n(i \mid \mathcal{C}) \]

• not allowing for all alternatives can lead to inconsistencies
  – well known in choice modeling
  – have never seen this in OD matrix estimation
Things can go wrong in the simplest case

- scenario:
  - peak hour demand of 1500 exceeds each route alone
  - congestion builds up *upstream* of the diverge
Things can go wrong in the simplest case

- scenario:
  - peak hour demand of 1500 exceeds each route alone
  - congestion builds up *upstream* of the diverge
- best-response choice set generation *never* finds the detour
- even with a stochastic choice model, *no one* takes the detour
- effect on OD/path flow estimation when using random slacks?!

\[ \text{cap} = 2000 \quad \text{cap} = 1000 \quad \text{cap} = 2000 \]

\[ \text{cap} = 1000 \]
Metropolis-Hastings sampling of paths (1/2)

• approach
  – give every path \( i \in C \) a weight \( b(i) > 0 \)
  – sampling probability \( q(i) \) shall be \( \propto b(i) \)

• direct sampling from \( q(i) \) requires path enumeration

\[
q(i) = \frac{b(i)}{\sum_{j \in C} b(j)}
\]

• MH does the job based on pair-wise comparisons only

\[
\frac{q(i)}{q(j)} = \frac{b(i)}{b(j)}
\]
Metropolis-Hastings sampling of paths (2/2)

[[movie]]^2

^2Flötteröd & Bierlaire (submitted).
transp-or.epfl.ch/documents/technicalReports/FloeBier11.pdf
Outline

Relevance: truthful modeling of uncertainty

Doability: avoiding to drown in details

Summary
Where this is going

• this talk suggests to simulate and calibrate more “details”
• here: an idea of how a “gradual enrichment” is possible
  – first, introduce disaggregation without additional modeling
  – second, exploit the resulting structure whenever convenient
Example: dynamic OD matrix estimation

- well-known to be utterly underspecified
- “macroscopic” approaches to deal with this:
  - non-negativity constraints
  - stay close to an (arbitrary) prior
  - assume (linear) dynamics
Example: dynamic OD matrix estimation

- well-known to be utterly underspecified
- “macroscopic” approaches to deal with this:
  - non-negativity constraints
  - stay close to an (arbitrary) prior
  - assume (linear) dynamics

- problems arise when facing:
  - rigorous mass conservation
  - truthful representation of demand fluctuations
  - more than a handful of commodities
Autoregressive OD matrix dynamics

- autoregressive model for OD flows (simplified):

\[ x_s(k + 1) = \sum_r a_{rs}(k)x_r(k) + \varepsilon_s(k) \]
Autoregressive OD matrix dynamics

- autoregressive model for OD flows (simplified):
  \[ x_s(k + 1) = \sum_r a_{rs}(k)x_r(k) + \varepsilon_s(k) \]

- expected OD flows are sums of choice probabilities:
  \[ \sum_n P_n(s|k + 1) = \sum_r a_{rs}(k) \sum_n P_n(r|k) \]
Autoregressive OD matrix dynamics

• autoregressive model for OD flows (simplified):

\[ x_s(k + 1) = \sum_r a_{rs}(k)x_r(k) + \varepsilon_s(k) \]

• expected OD flows are sums of choice probabilities:

\[ \sum_n P_n(s|k + 1) = \sum_r a_{rs}(k) \sum_n P_n(r|k) \]

• looking back at the original problem:

\[ \iff \forall n : P_n(s|k + 1) = \sum_r a_{rs}(k)P_n(r|k) \]
Really more complex than the AR model?

- Markovian trip making dynamics at individual level
- truthful representation of original AR model
- but...
Added value of traveler disaggregation

- constraints become simple in the disaggregate approach:
  - rigorous mass conservation
  - truthful representation of demand fluctuations
  - more than a handful of commodities

- in addition, one can add any behavioral model of trip chaining
- in this particular example, all of this is already possible\(^3\)

Outline

Relevance: truthful modeling of uncertainty

Doability: avoiding to drown in details

Summary
Summary

- **disaggregate simulation & calibration modeling**
  - capture uncertainty where it occurs
  - contribute to unbiased calibration

- **model complexity does not necessarily explode**
  - in the 1st instance, only add physically existing structure
  - in the 2nd instance, add more complex model structure

- **(very subjective) conclusion**
  - calibration models benefit from increased disaggregation
  - possible without jumping right on activity-based models