WORKER-JOB MATCHING AND SHADOW-PRICING IN ACTIVITY-BASED MODELS

JOHN GIBB

Senior Transportation Engineer

DKS Associates

jag@dksassociates.com

MODELING MOBILITY CONFERENCE 2025

SHAPING A SMARTER TRANSPORTATION EXPERIENCE™

DKSASSOCIATES.COM

AN EMPLOYEE-OWNED COMPANY

WORK LOCATION MODEL

Agents: Employed persons

Choice set: Spatial units, employment given

Choice model + market model

Logit + maximum entropy is

$$p_i(j) = \frac{\exp(\boldsymbol{\beta x}_{ij} + \alpha_j)}{\sum_{\tilde{j} \in \{J\}} \exp(\boldsymbol{\beta x}_{i\tilde{j}} + \alpha_{\tilde{j}})}$$

α_j = "shadow prices"

Same for all agents

Solution: $\alpha_j \leftarrow \alpha_j + \ln\left(\frac{w_j}{\sum_i p_i(j)}\right)$ for all j

given

predicted

WHY REVISIT SHADOW-PRICES?

A solved problem???

Disaggregate individuals	Long **runtimes**
Single random outcomes	**Noise:** $n_j \sim \mathrm{Poisson}(\sum_i p_i(j))$
Unchosen alternatives	Must adapt to not **divide by zero**
Adaptations w/ **limited & unpublished investigation**	Convergence "**bottoms out**"

Can we solve better? Faster?

“CLEARINGHOUSE” MARKET MODELS
≠ MAX ENTROPY

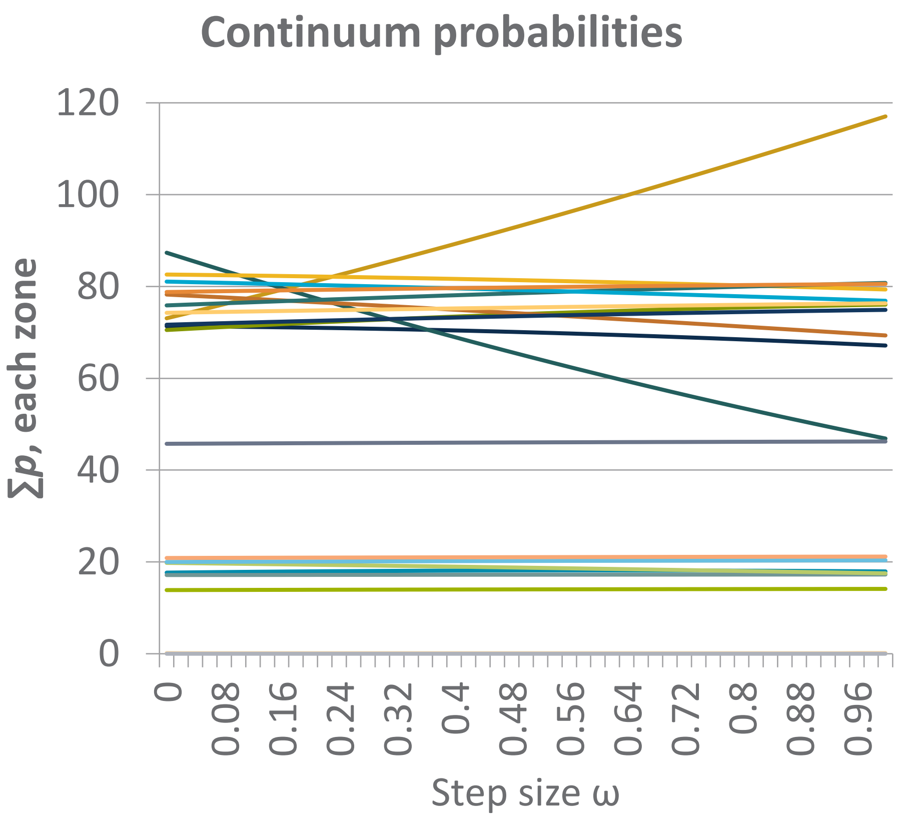
Ordered choice = *Serial Dictatorship*

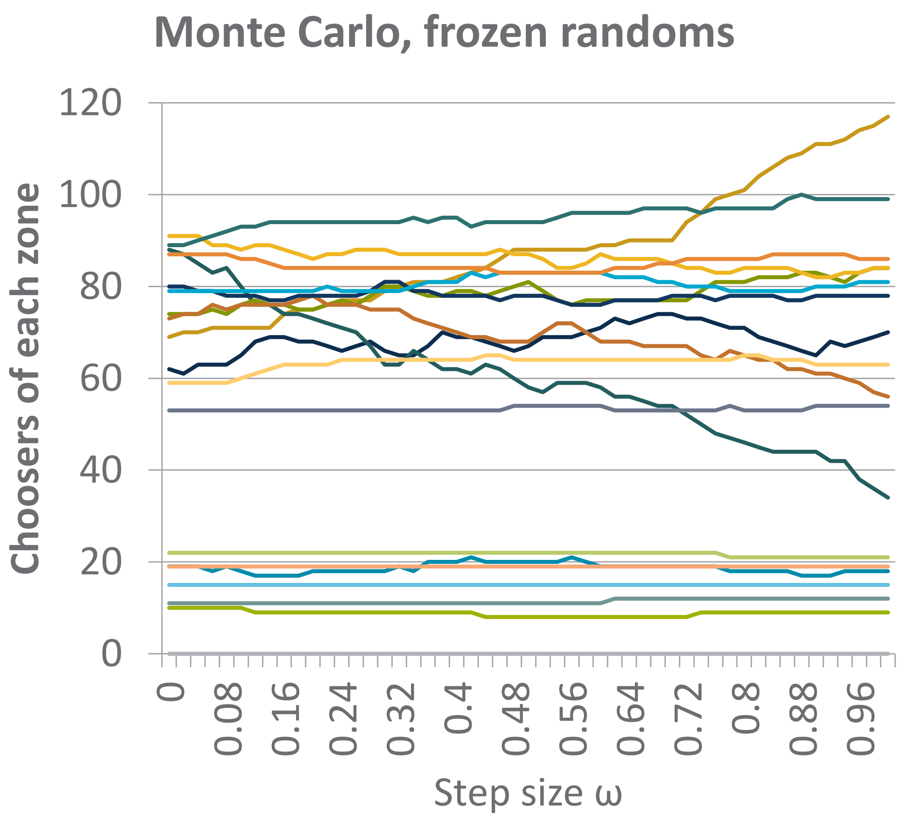
ActivitySim ≈ *Rank Maximal*

DEMO – NOISE

1000 persons, 20 alternatives

Add an ASC vector in small steps





REGIONAL MODEL *EXPERIMENTS*

MODEL FOR EXPERIMENTS

Rough emulation of Sacramento regional model work location choice

- 1,046,000 working persons
- 15,900 “parcel” groups having employment
  - Grouped to avoid < 10 if possible
- 100 sampled alternatives per person

ADJUSTMENT FORMULAS IN USE

Handle **zero outcomes**

Many **dampen**

CTRAMP: $\alpha \leftarrow \begin{cases} \alpha + \omega \ln\left(\frac{w}{n}\right) \text{ if } n > 0 \\ \alpha \text{ otherwise} \end{cases}$

Daysim: $\alpha \leftarrow \alpha + \ln\left(\frac{w \pm tol}{\max(n, 0.01)}\right)$, *tol* by diff or %

Truncate: $\alpha \leftarrow \alpha + \omega \ln\left(\frac{w}{\max(n, \delta)}\right)$, δ =0.5, 1, ...?

ITERATION HISTORY OF SOME LOCATIONS WITH **ZERO ITERATES: CTRAMP**

	Iteration ->							
Jobs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
16	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0
29	0	1	16	26	24	27	28	27
15	0	1	9	9	15	13	13	14
39	0	0	0	1	9	22	27	32
16	0	0	2	1	17	18	19	18
10	0	0	1	7	7	6	7	9
10	0	1	7	8	8	9	9	8
40	0	2	19	33	34	37	35	39

ITERATION HISTORY OF SOME LOCATIONS WITH **ZERO ITERATES: DAYSIM**

	Iteration ->							
Jobs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
30	0	0	0	0	1687	161	36	32
6	2	0	1	2	2	2	1	1
10	0	0	0	0	47	9	6	20
5	1	0	1	0	0	1	1	0
14	0	371	39	19	10	15	14	13
11	0	0	0	0	73	10	10	10
6	6	1	0	0	1	0	1	1
7	0	1	1	1	3	3	2	1
11	0	0	0	0	45	14	9	14

NEW ADJUSTMENT FORMULAS

Size-based dampening (by zone size)

S1: $$\alpha \leftarrow \alpha + \omega \ln\left(\frac{w+1}{n+1}\right)$$

S2: $$\alpha \leftarrow \alpha + \ln\left(\frac{w+\delta}{n+\delta}\right)$$

S3: $$\alpha \leftarrow \alpha + \ln\left(\frac{w+\theta w+\delta}{n+\theta w+\delta}\right)$$

Typical: δ=1 or experimental, increase later; ω=1, decrease gradually

Difference-based dampening (by error difference)

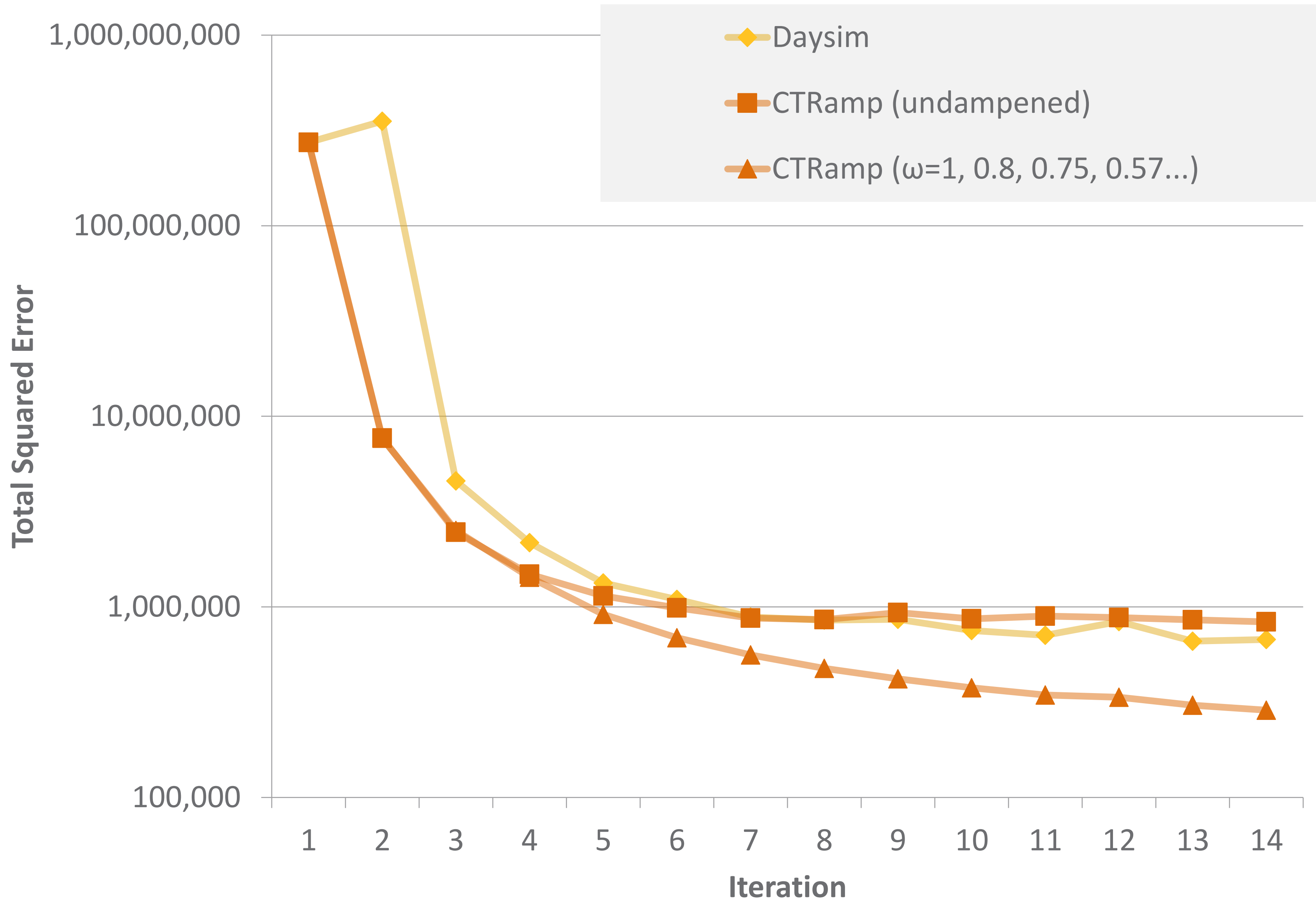
D1: $$\alpha \leftarrow \alpha + \ln\left(\frac{w}{n+(w-n)\frac{\delta}{\delta+|w-n|}}\right)$$

D2: $$\alpha \leftarrow \alpha + \ln\left(\frac{w}{n+(w-n)\frac{\delta^2}{\delta^2+(w-n)^2}}\right)$$

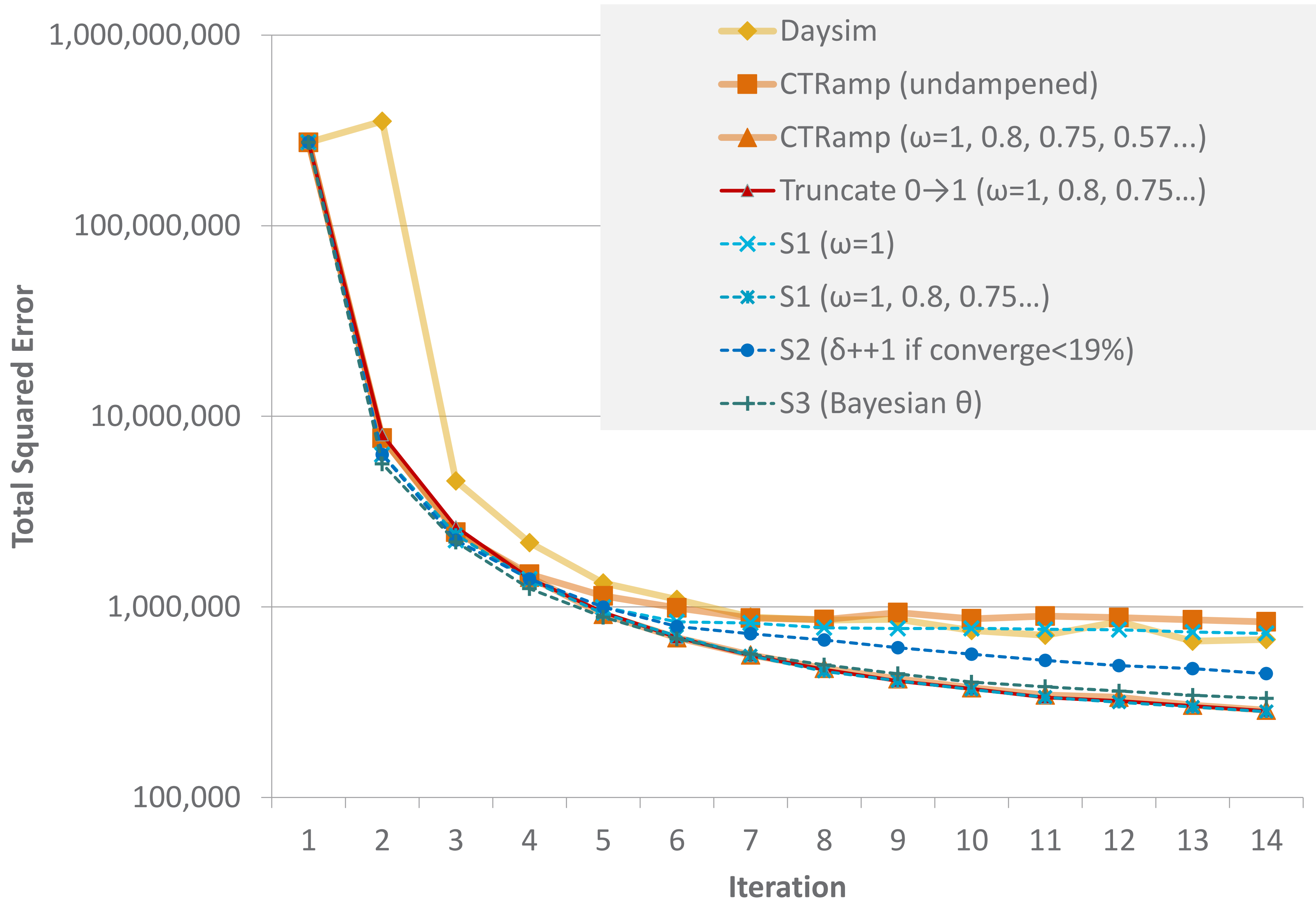
Toward *w*…

…by a fraction

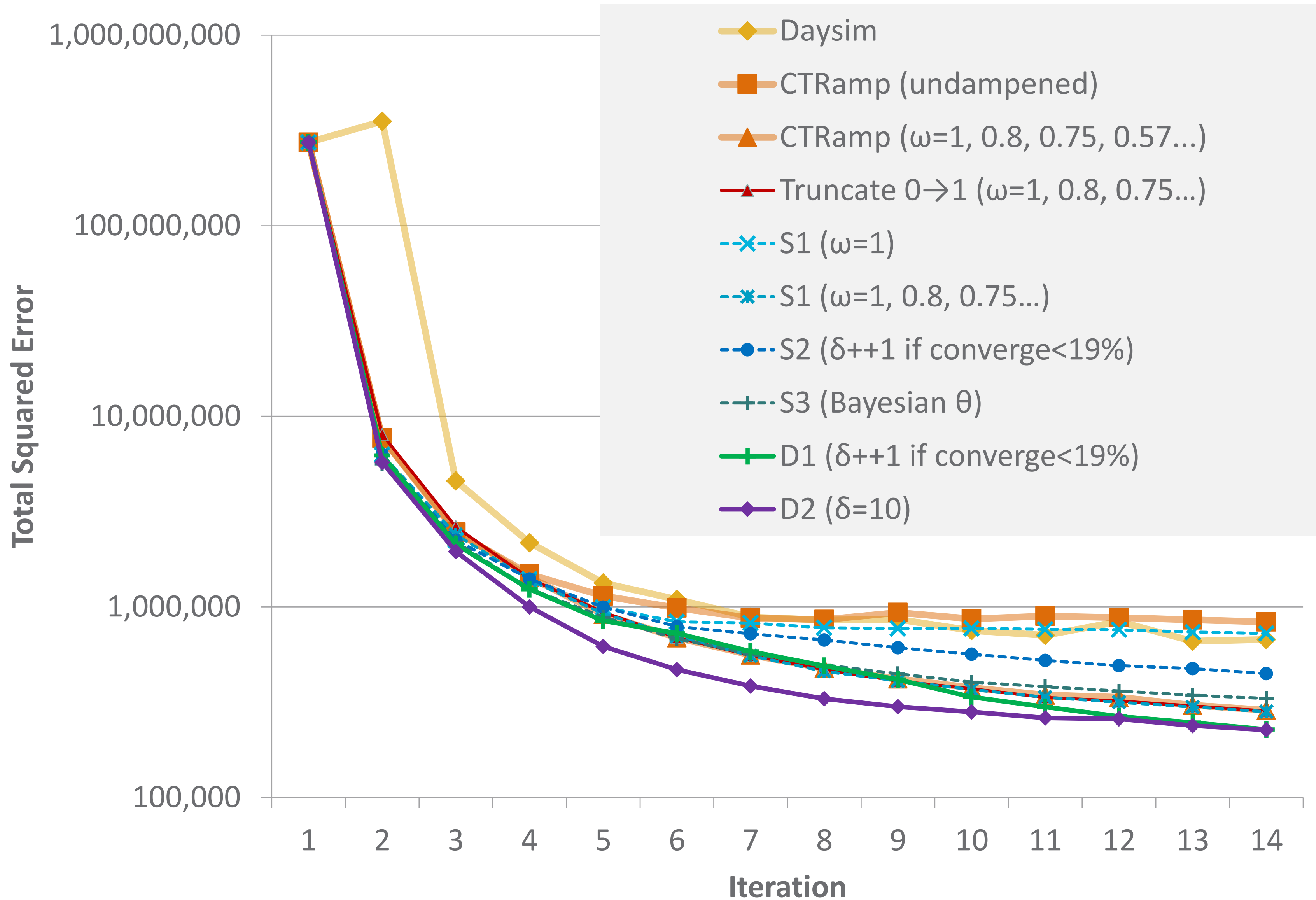
CONVERGENCE: BASIC FORMULAS



COMPARE SIZE-DAMPENING



COMPARE DIFFERENCE-DAMPENING



MONTE CARLO *IS NOT THE ONLY CASINO*

APPLICATION METHOD 2: FROZEN RANDOM UTILITIES

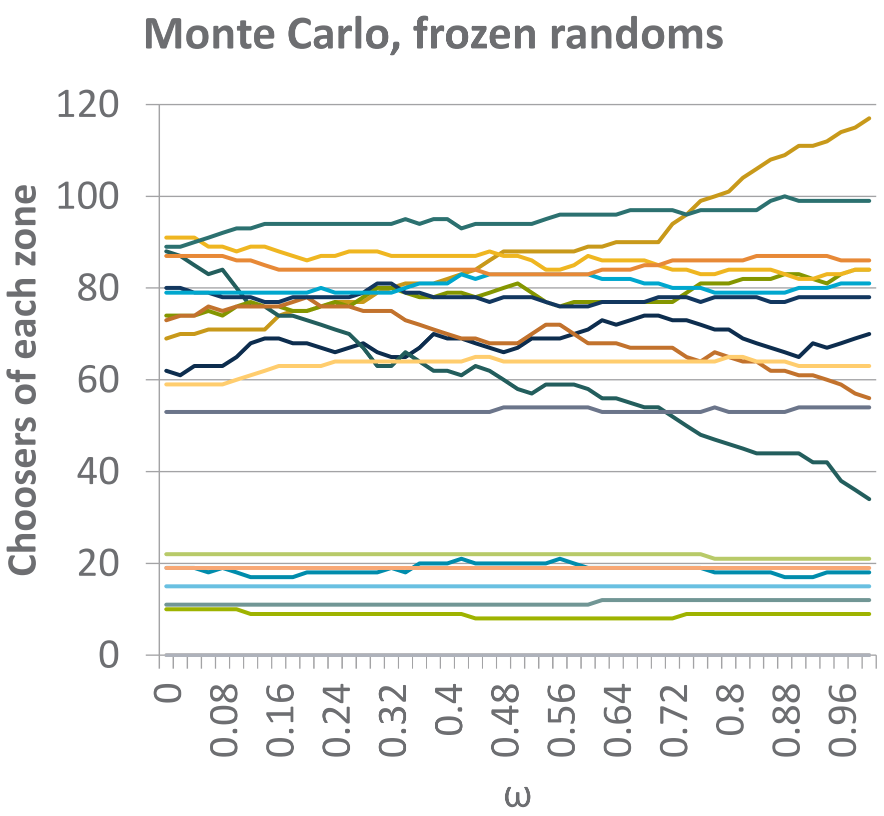
Monte Carlo → **counterintuitive & excessive choice switching** (Zill & Veitch, 2022)

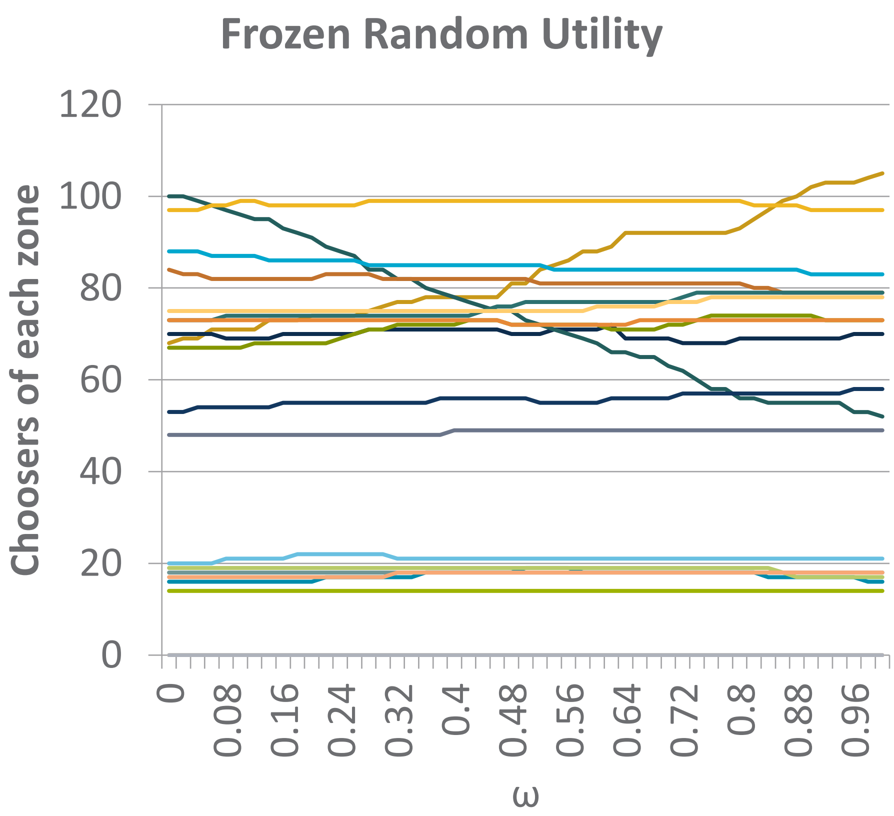
Solution: draw Gumbel random utilities

$$\epsilon_{ij} = -\ln(-\ln(random \in (0,1)))$$

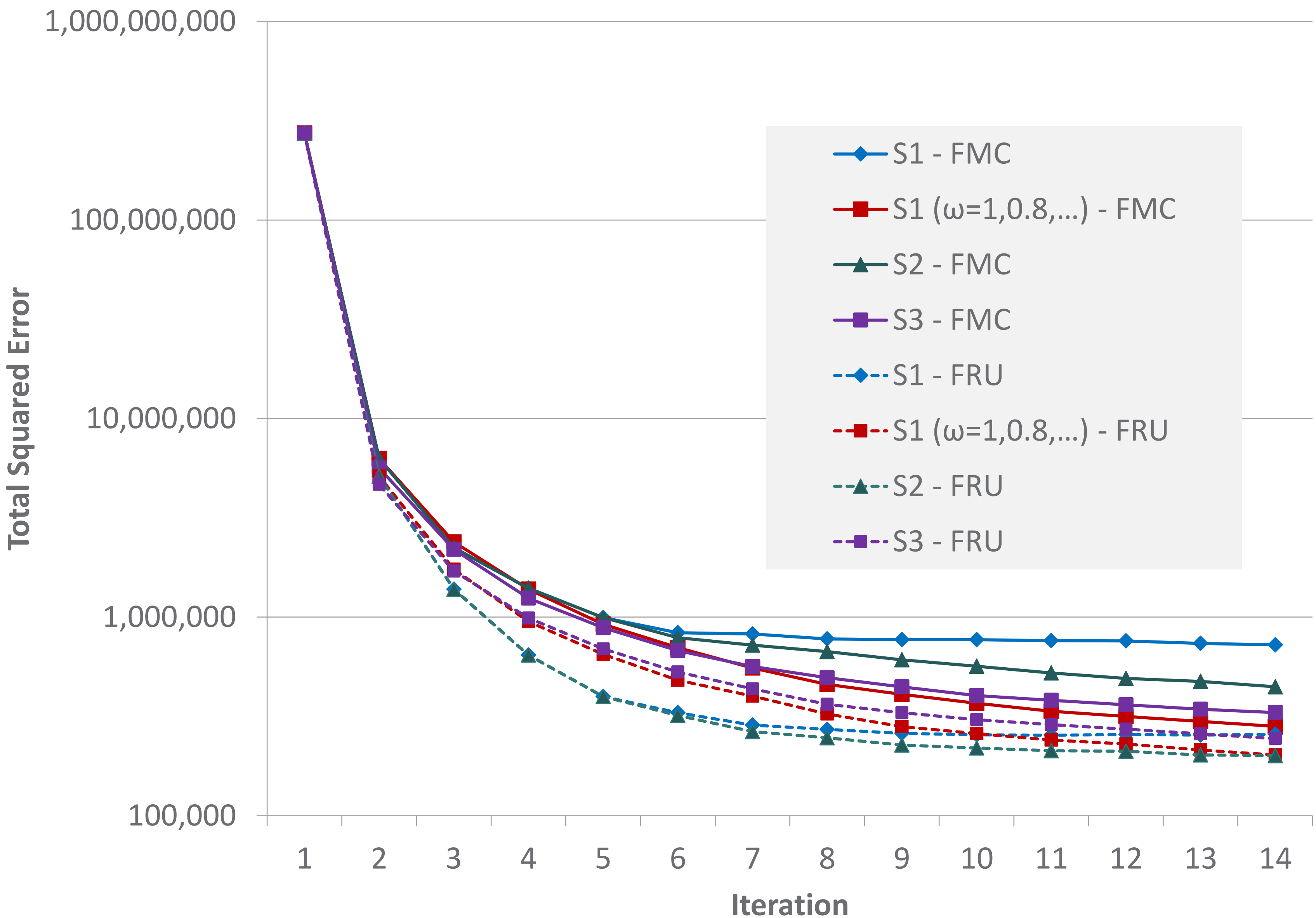
DEMO – NOISE

Same 20-zone as before

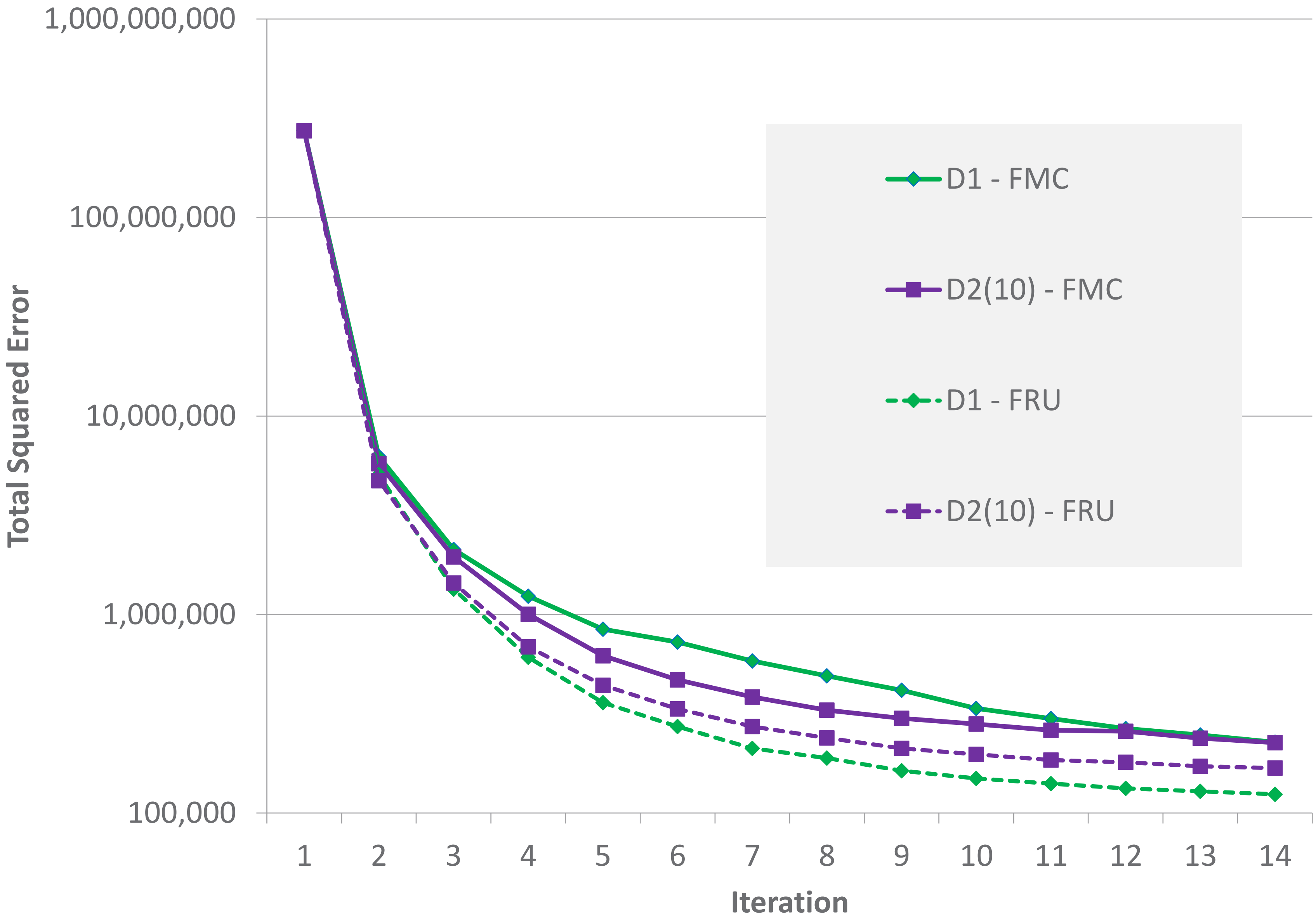




CONVERGENCE: FMC vs FRU, *S* forms



CONVERGENCE: FMC vs FRU, *D* FORMS

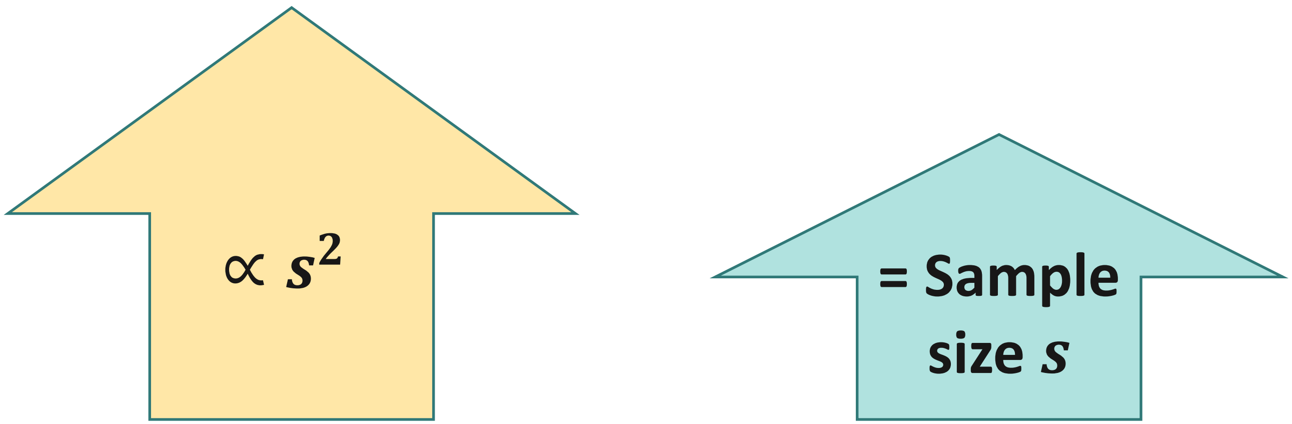


SHADOW-PRICING WITH *SAMPLES* OF THE POPULATION?

ERROR DECOMPOSITION

Sample squared error

= systematic error + Poisson variance

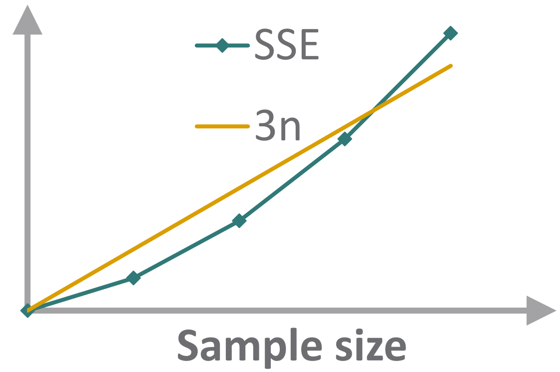


Signal *Noise*

Big-enough sample if “signal” >> “noise”

AGENT SAMPLING METHOD

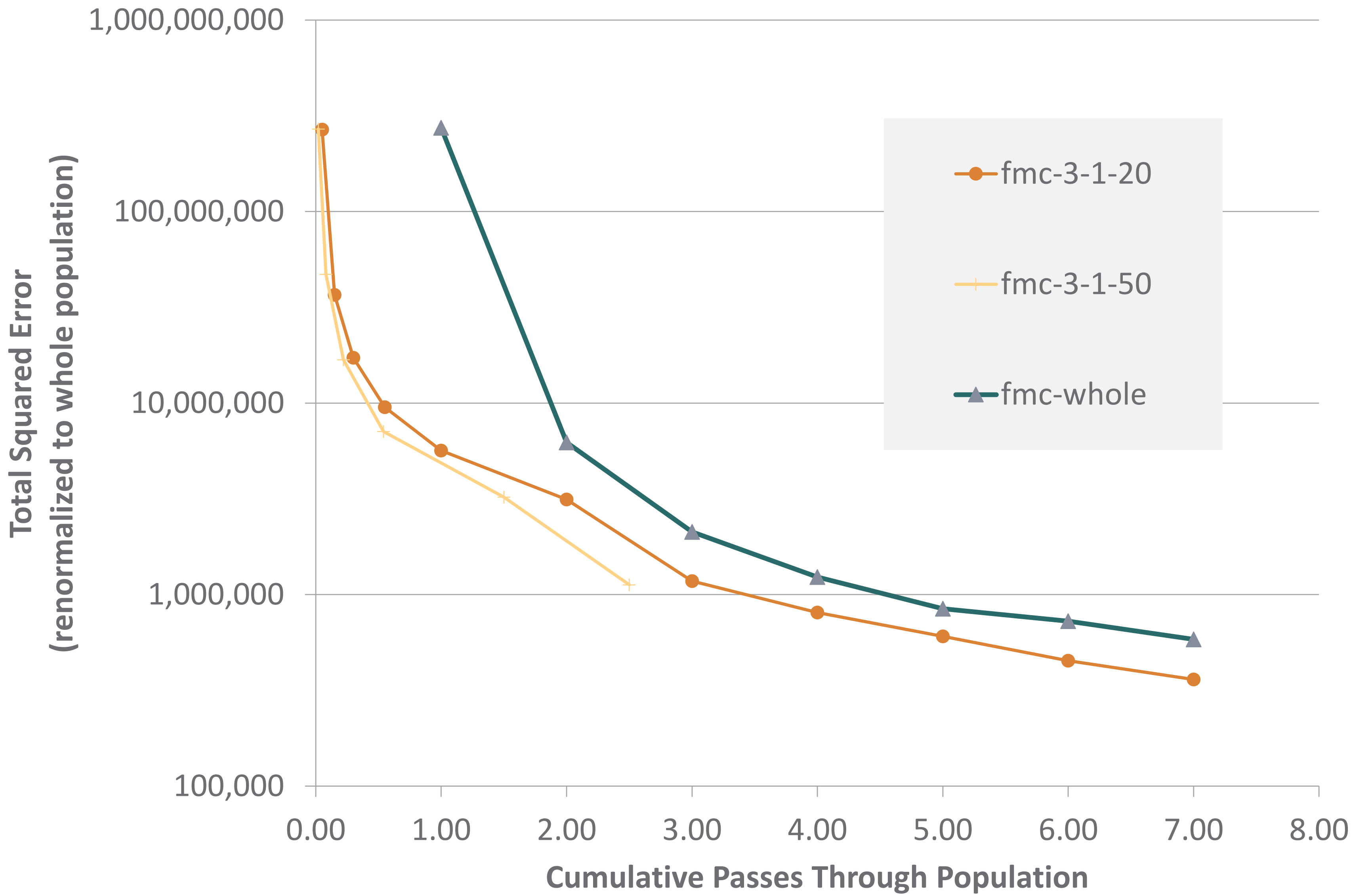
Batch 1	Sample 1
Batch 2	
Batch 3	Sample 2
Batch 4	
Batch 5	
Batch 6	
Batch 7	Sample 3
Batch 8	
Batch 9	
Batch 10	
Batch 11	
Batch 12	Begin Sample 4?
Batch 13	
Batch 14	
Batch 15	
Batch 16	
Batch 17	



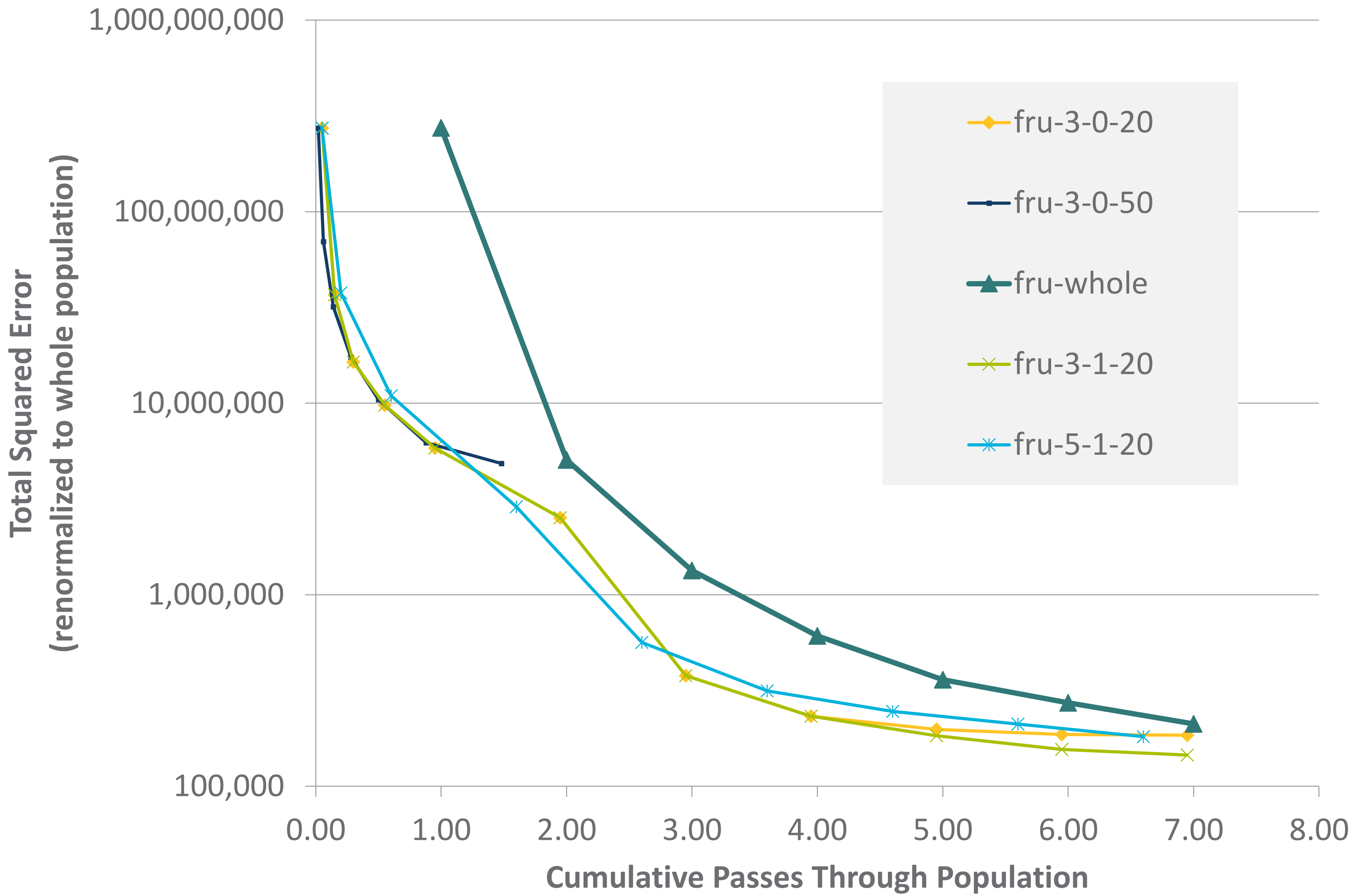
IF *samp sq err > sample size* • (3 or so),
OR sample = full population,
THEN Adjust SPs, Start new sample
ELSE keep on with current sample

Return to beginning

GROWING SAMPLE TESTS – FMC



GROWING SAMPLE TESTS – FRU



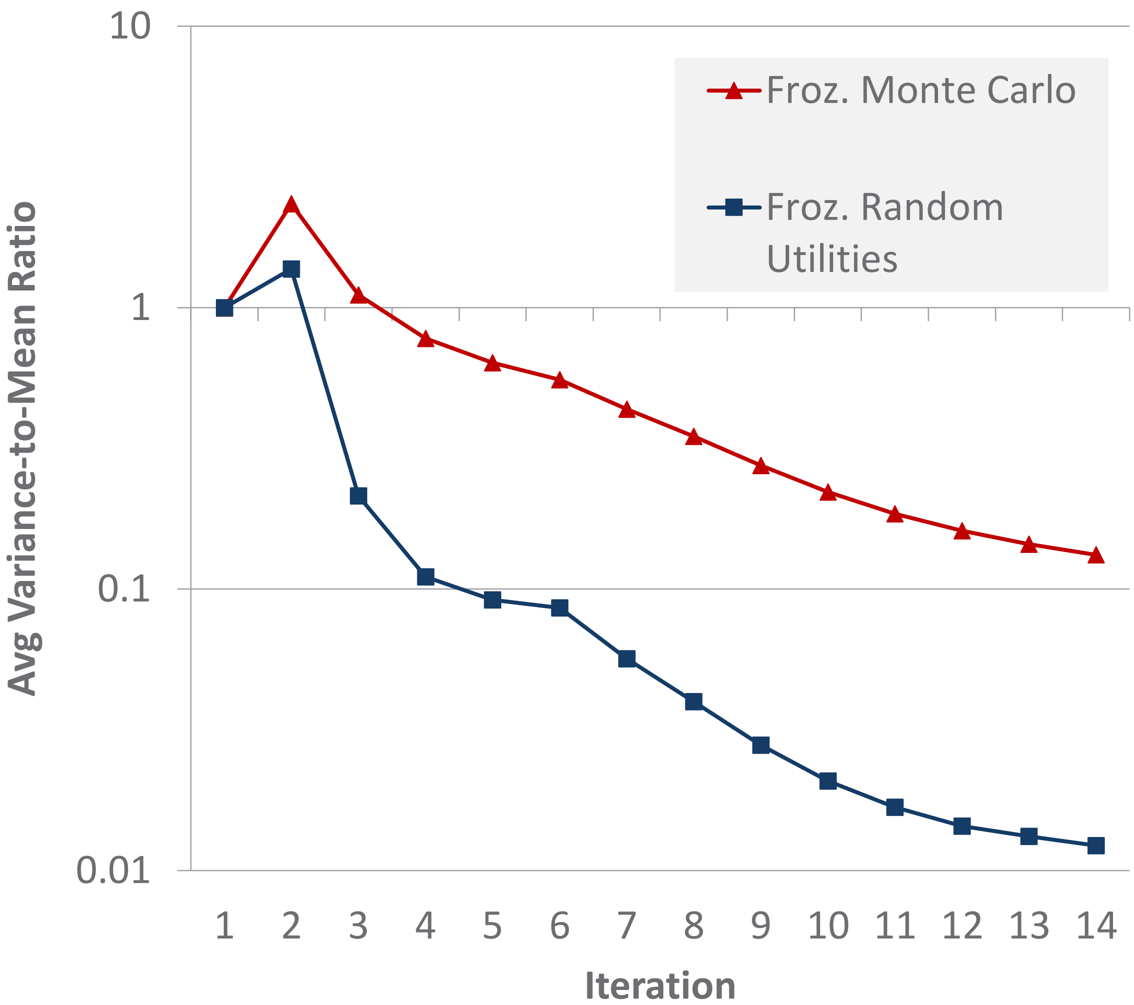
STOCHASTIC VARIATION

STOCHASTIC VARIATION

10 runs of selected models (dampened-difference)

Variance is per location

Same inputs except **random numbers** (seed)



WRAP-UP

CONCLUSIONS

Use **better SP adjustment formulas**

esp. difference-based dampening

Clearinghouse methods are not substitutes

Pursue **frozen random utilities** instead of Monte Carlo

Shave runtime with **agent sampling** runtime method

EXTRA SLIDES FOLLOW

DIAGNOSTICS FOR BOTTOMING OUT

Especially squared error > Poisson error

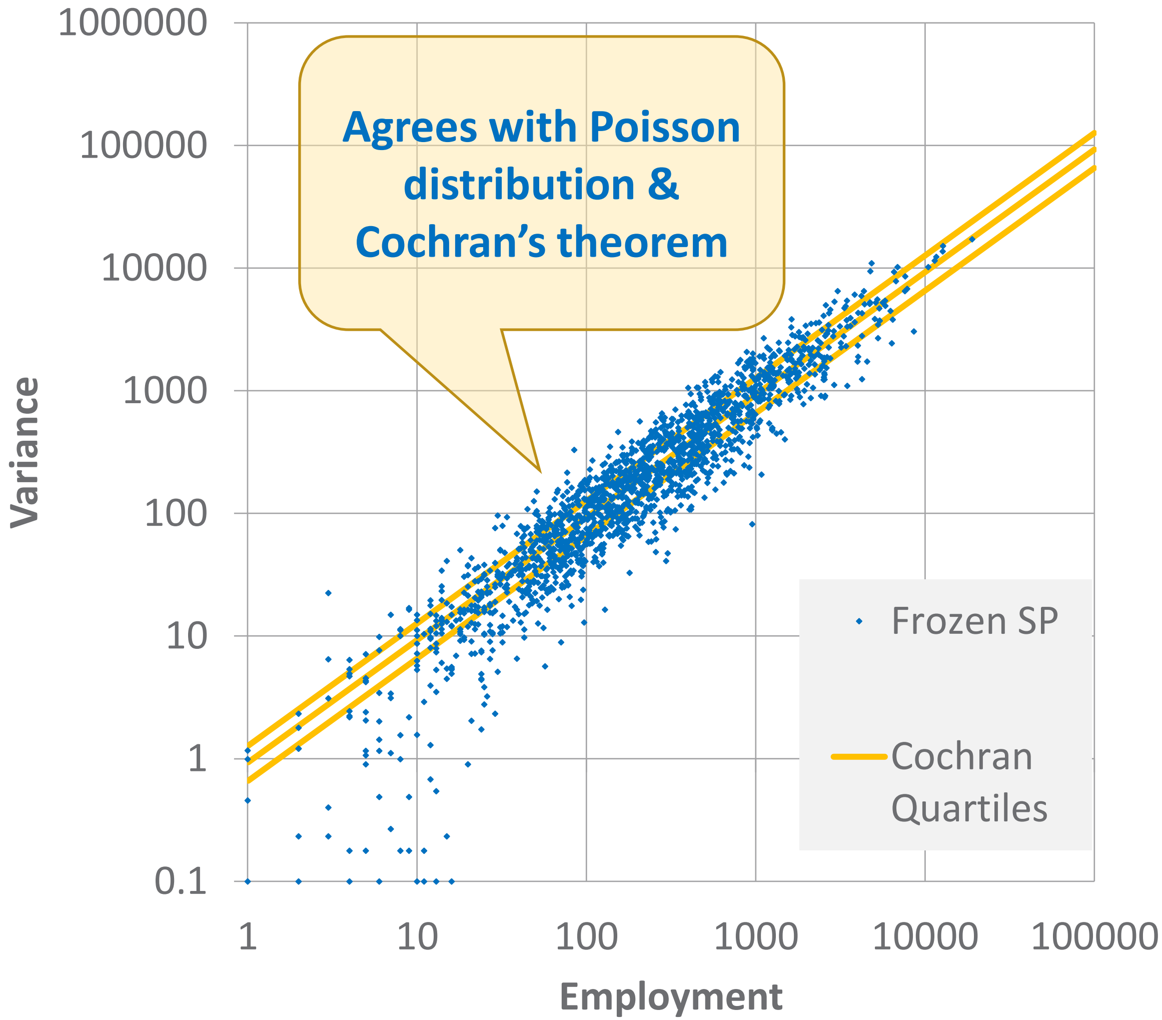
- Locate outliers, “pockets of resistance”, isolated areas with supply-demand imbalance, SPs diverge with little change in choice, or SP >> ln(num of sample alts)
  - Data errors? IX-XI problems?
  - Do more sampled alternatives help?
  - Try flatter sampling function, to include more longer-distance locations. (You can’t have conditional probability > 1 to make up for undersampling.)
- School choice may suffer local imbalances, data uncertainty. Matching forces excessive long commutes. Consider *soft constraints*.

MORE ON

STOCHASTIC VARIATION

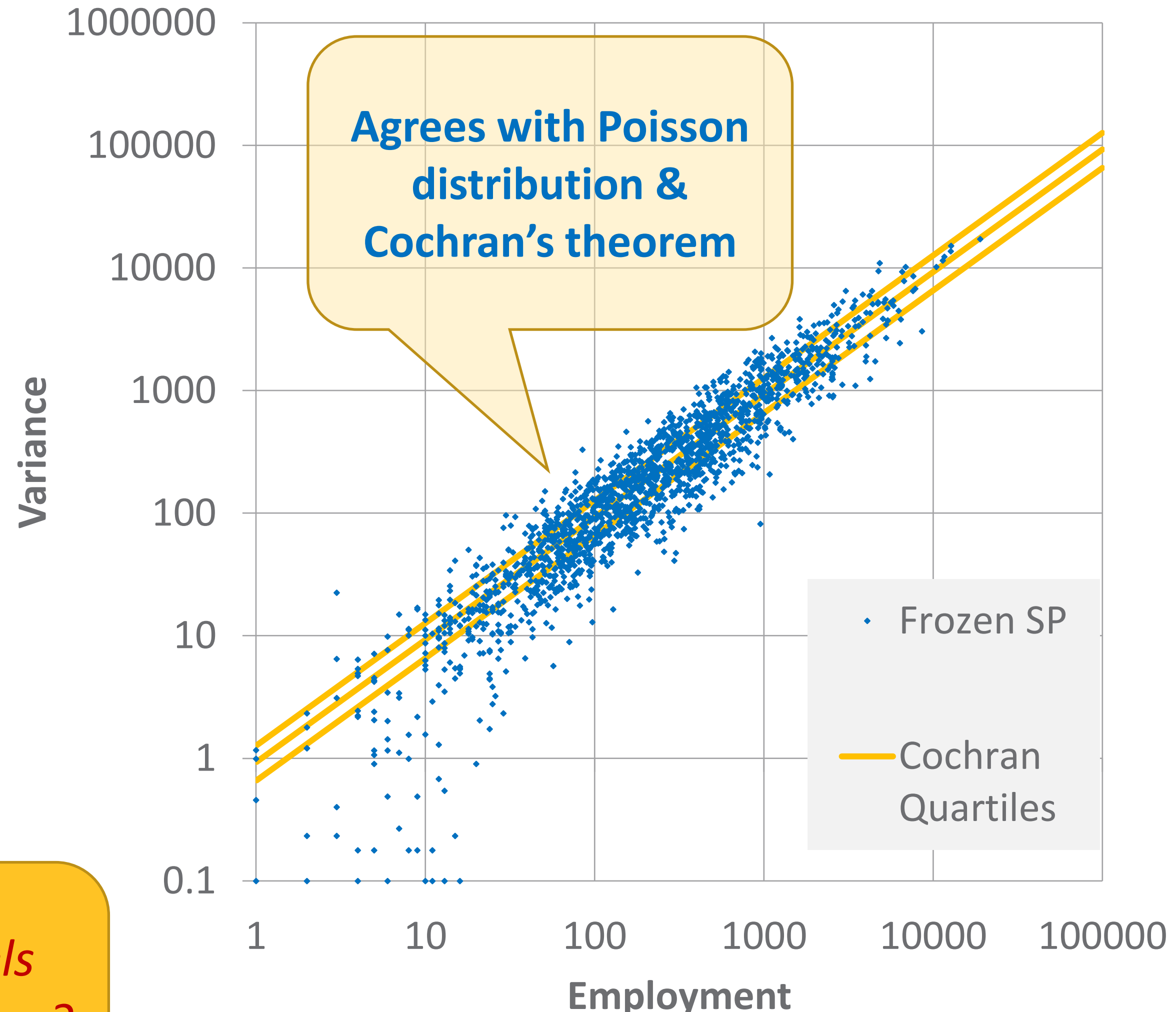
VARIANCE OF WORKERS BY TAZ (1)

- 10 runs of Sacramento ABM alone
- All inputs identical, including SPs
- Only the random seeds vary between runs



VARIANCE OF WORKERS BY TAZ (1)

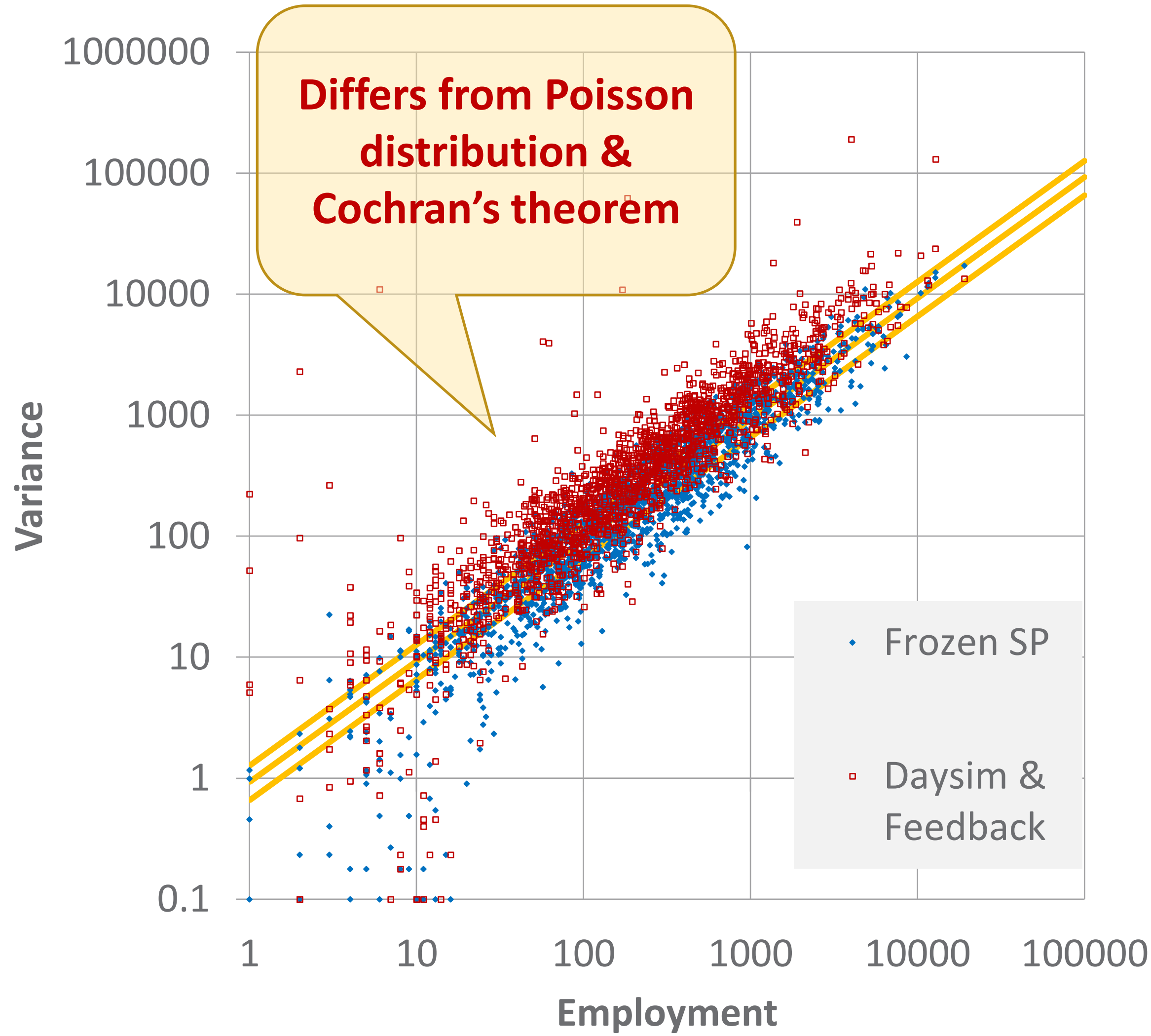
- 10 runs of Sacramento ABM alone
- All inputs identical, including SPs
- Only the random seeds vary between runs



How should full models go, each with SP iteration? More? … Less?

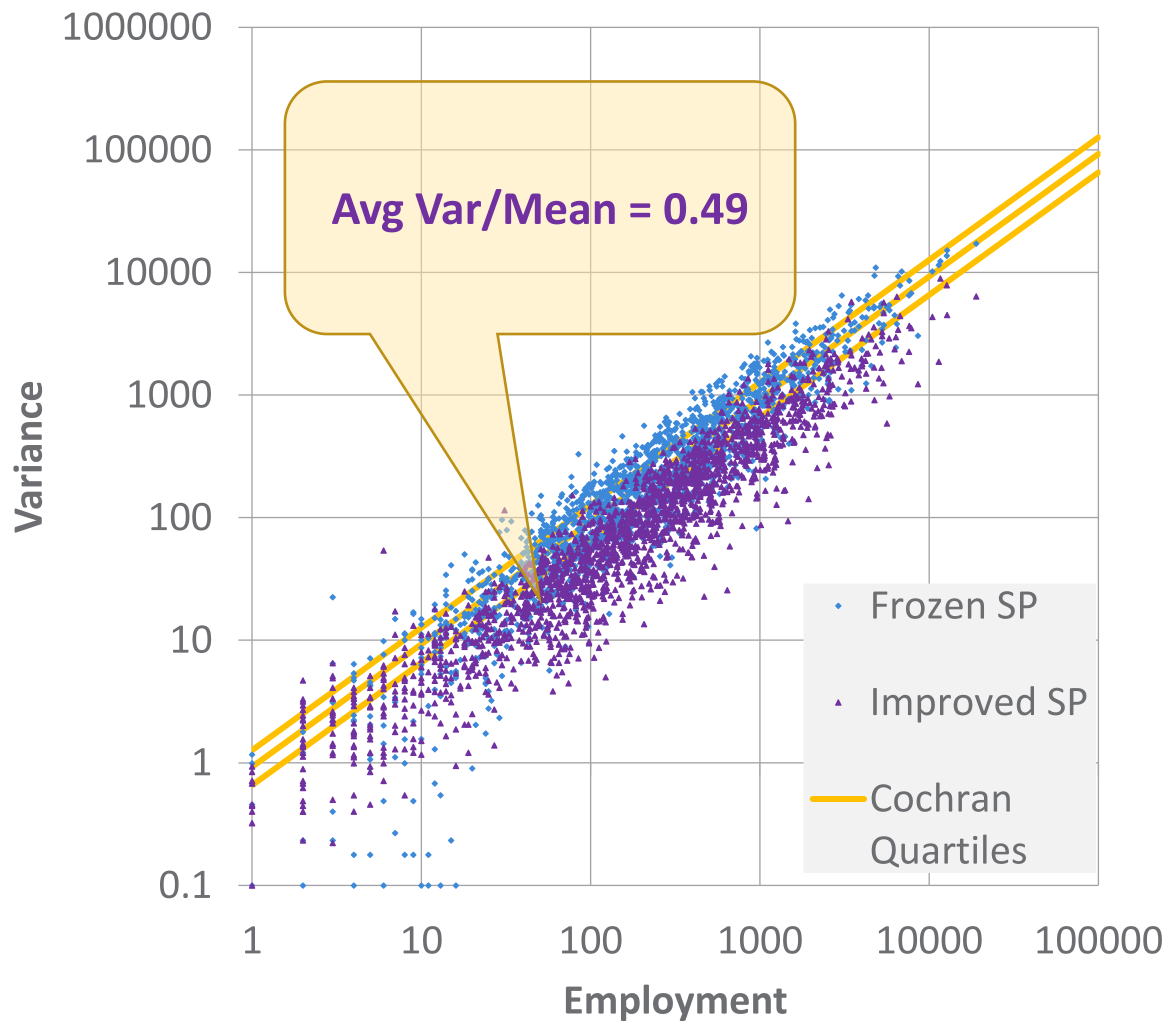
VARIANCE OF WORKERS BY TAZ (2)

- 10 runs of entire Sacramento TDM w/ SP iteration
- Initial inputs identical
- Only the random seeds vary between runs



VARIANCE OF WORKERS BY TAZ (3)

- 10 runs of entire Sacramento TDM w/ **new SP formula**
- Initial inputs identical
- Only the random seeds vary between runs



MORE ON

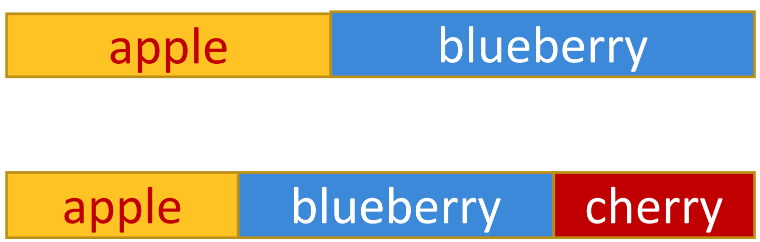
MONTE CARLO vs RANDOM UTILITIES

MORGANBESSER'S DESSERT CHOICE

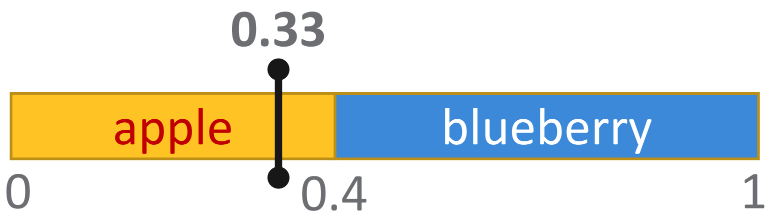
Sidney Morganbesser (1921-2004), American philosopher, social theorist

apple	blueberry

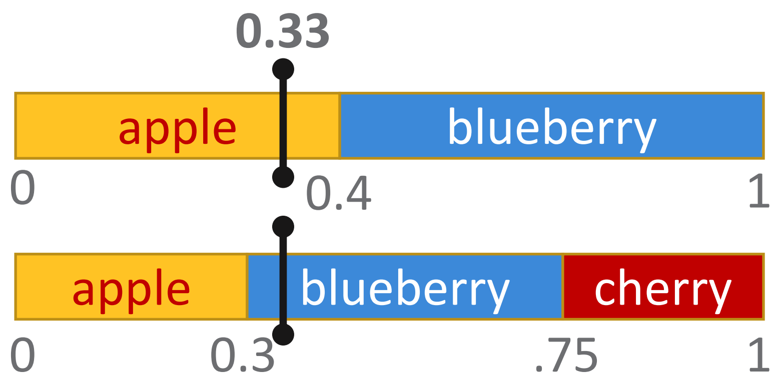
MORGANBESSER’S DESSERT CHOICE



MONTE CARLO AND MORGANBESSER’S DESSERT CHOICE



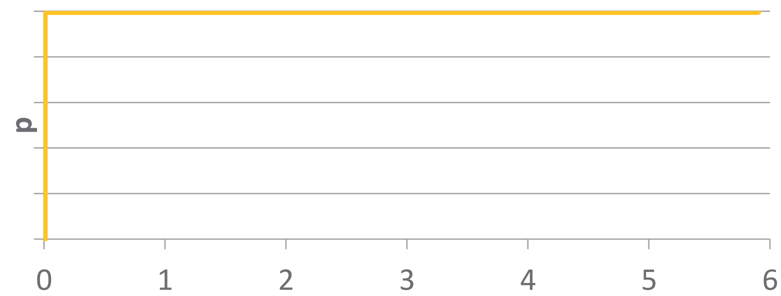
MONTE CARLO AND MORGANBESSER'S DESSERT CHOICE



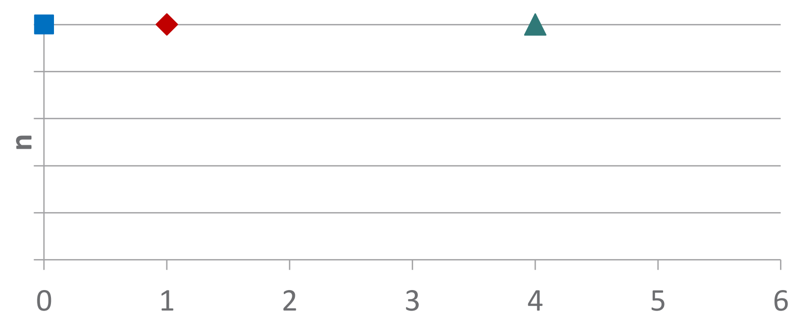
BAYESIAN PERSPECTIVE

A BAYESIAN VIEW

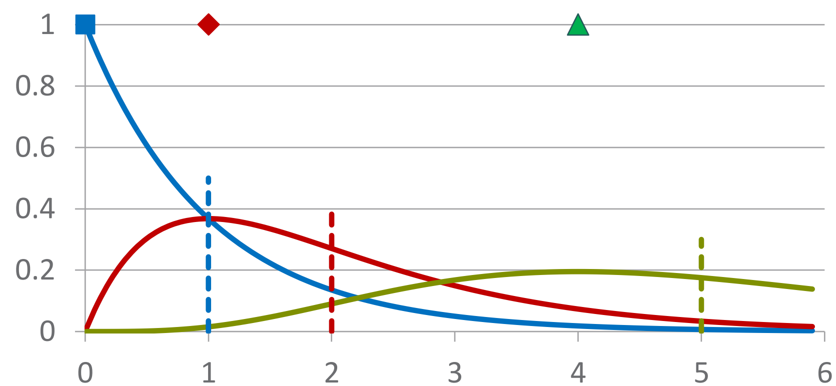
Uninformed prior



+ One Poisson-distributed observation



→ Posterior

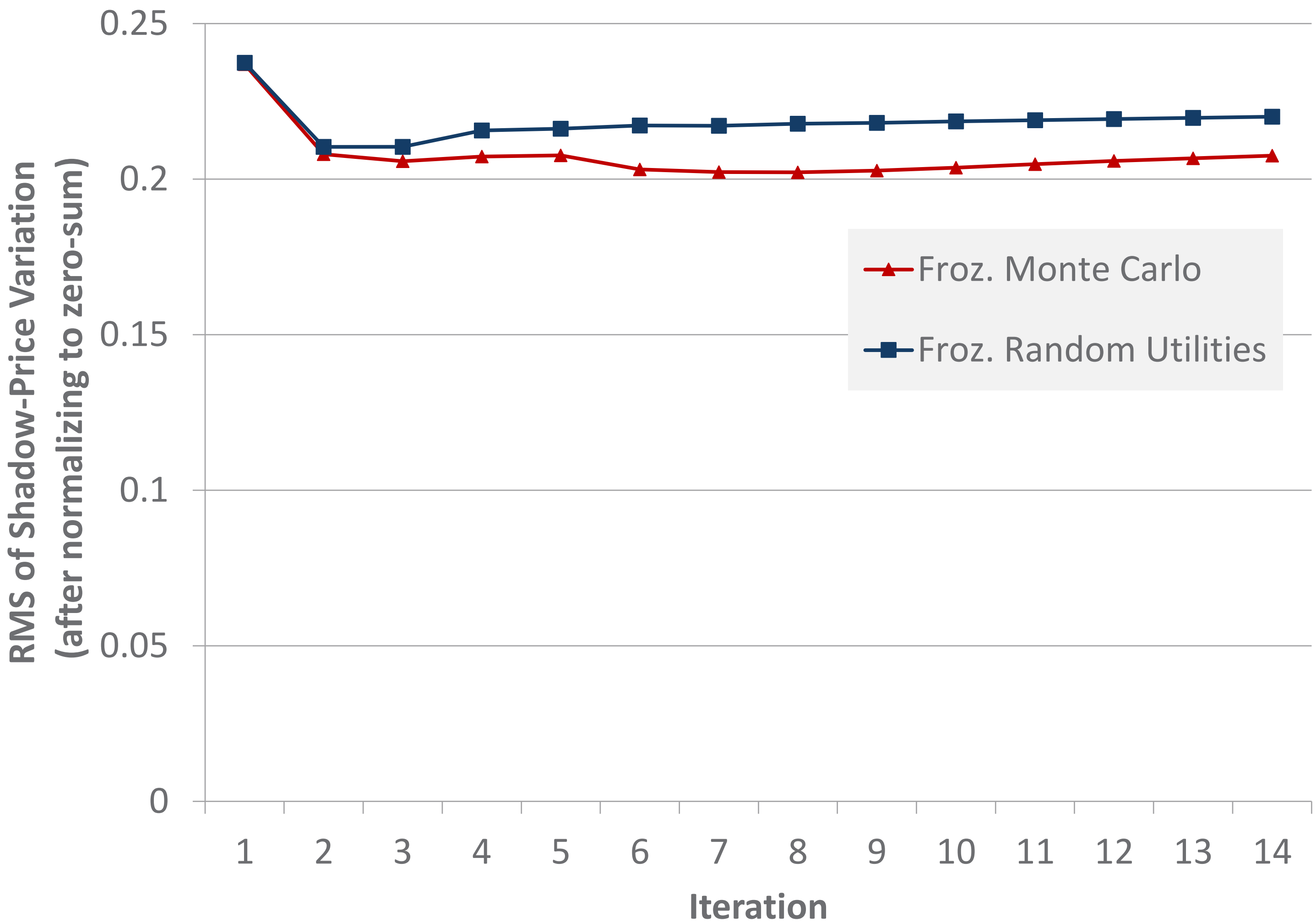


Mean = Obs + 1

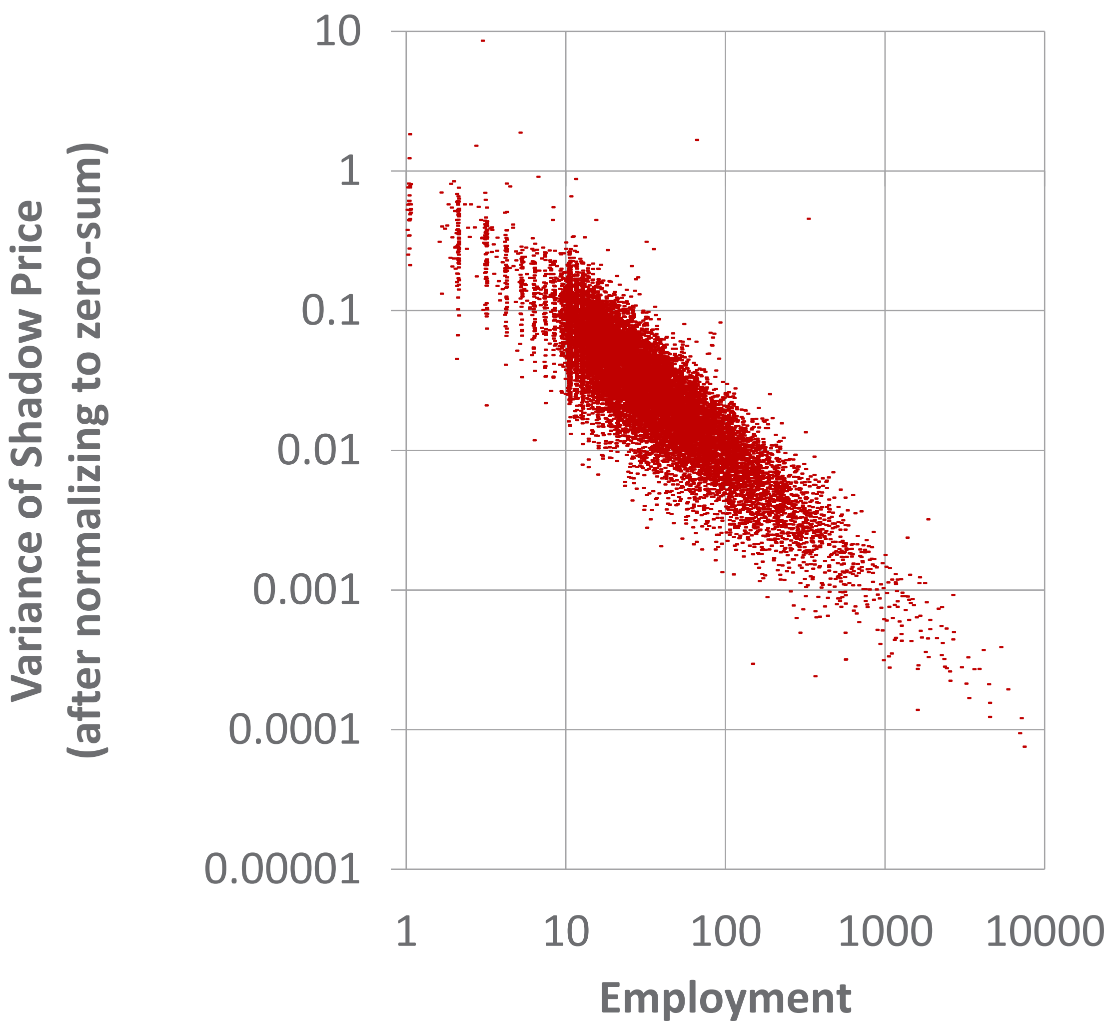
MORE ON

STOCHASTIC VARIATION OF SHADOW PRICES

STOCH. VARIATION OF SHADOW PRICES



STOCH. ERROR OF SHADOW-PRICES W.R.T. SIZE OF LOCATION



Result for 10 runs, each well-converged: 14th iteration of FRU using damped-diff

Shows “inverted” Poisson error in the well-converged shadow prices

WANT SHADOW-PRICE PRECISION??

Why?

- Dependent models
- User-benefits

How?

- Accumulate conditional probabilities instead of single outcomes
- Then neutralize shadow-prices: subtract the weighted average

APPLICATION METHOD 3: CONDITIONAL PROBABILITIES

Accumulate conditional probabilities, instead of single outcomes

100 alternatives have ≈ 1/50 the noise variance

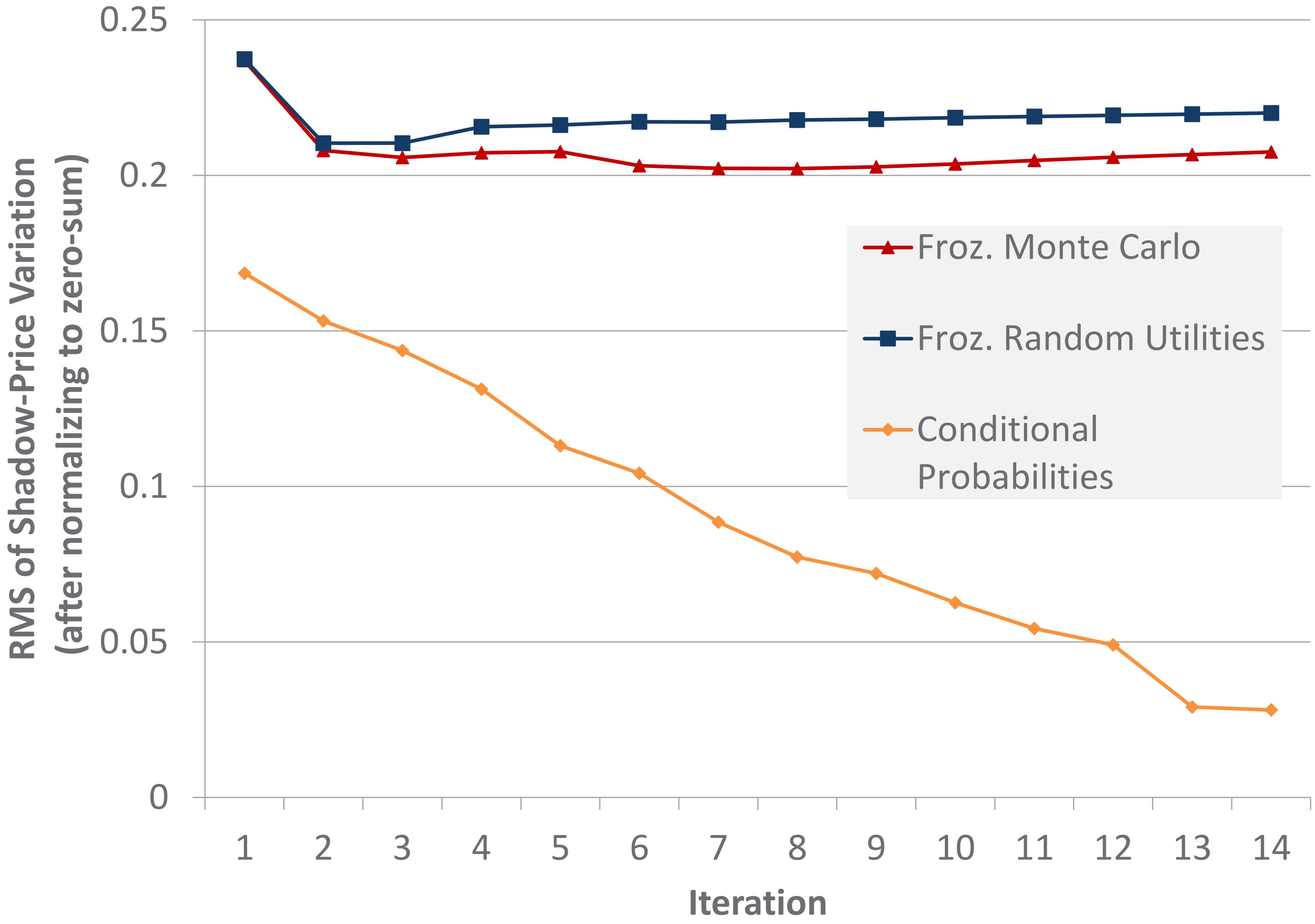
No frozen randoms – converge toward central limit

Can use samples of the population

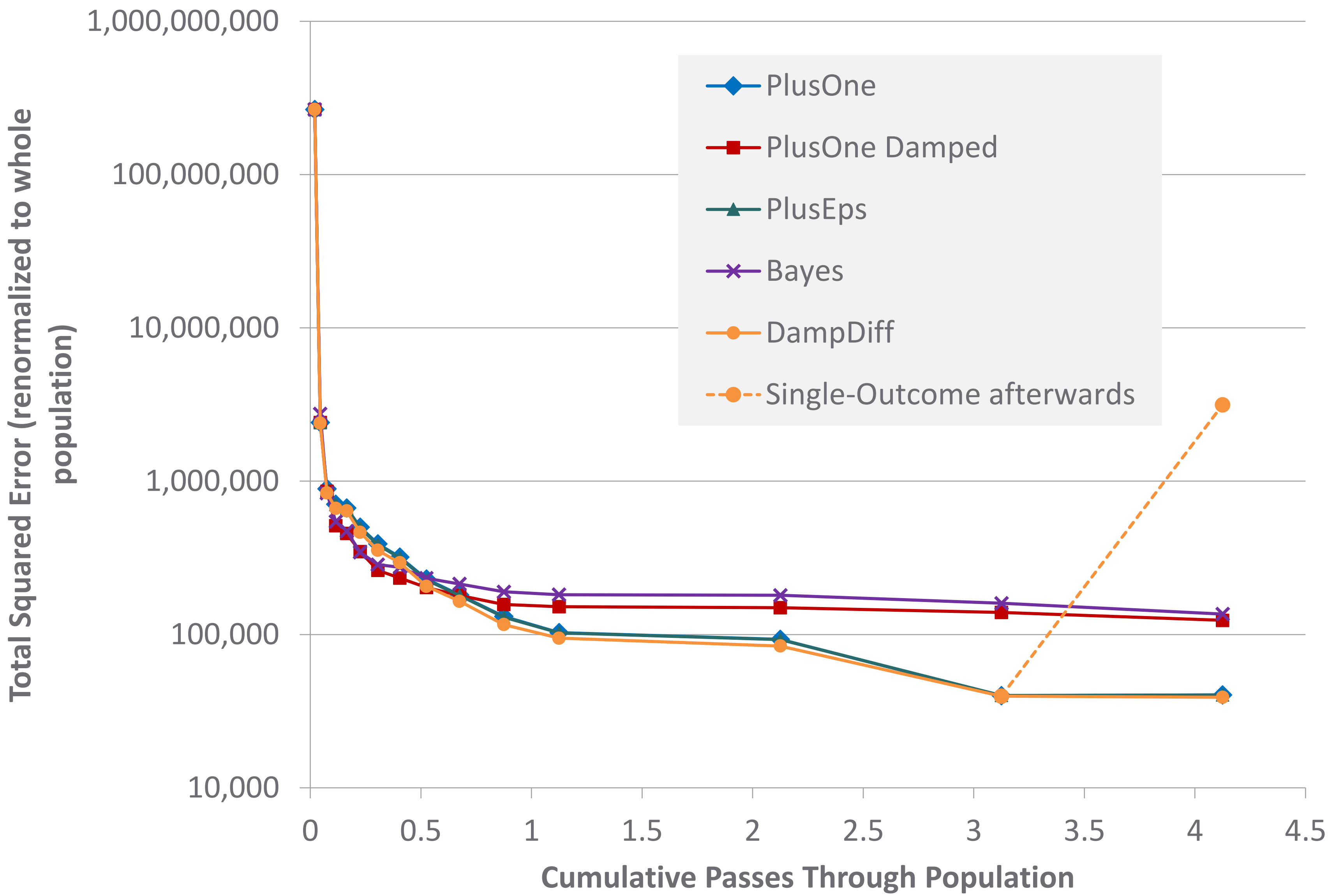
Need to draw single outcomes afterwards

- Or continue iterating with a single-outcome method

STOCH. VARIATION OF SHADOW PRICES



CONDITIONAL PROBS CONVERG'CE



MORE ON

AGENT SAMPLING

SIGNAL AND NOISE

- **In early iterations,**

  Squared error = Poisson variance + systematic error

- **From a sample of the population** (s out of N),

  Calculated Sample Sq'd Error $\equiv \sum_j \left(n_j - w_j \frac{s}{N}\right)^2$

  Sample Sq'd Error $\approx s + \left(\frac{s}{N}\right)^2 \left(\text{Popul. Sq}'\text{d Error} - N\right)$

- **As sample gets larger**, systematic error "signal" grows disproportionally over Poisson "noise".

  > Acceptance criteria tested: SSE>3s or SSE>5s

- **Later,** when acceptance criteria can't be met, process whole population

AGENT SAMPLING PROCEDURE

Select batches of sample agents (I used 1/20 or 1/50)

- **Run model** for everyone in the batch, accumulating to the current sample
- **Calculate** sample squared error (job targets scaled proportionally)
- **Test**: IF [sample squared error > sample size • (3 or so),
  AND cum. sample size > prev. sample size • (1.5 or so)]
  OR sample is the whole population, THEN
  - Update shadow prices
  - Reset current sample to empty

Repeat for the next batch

Ensure everyone's final choice uses final shadow prices

MORE ON

ENTROPY vs OTHER MARKET MODELS

SHADOW-PRICED LOGIT IS

- Maximum entropy • maximum total expected utility (logsums) • most-probable posterior • fair (same "prices" for all) • symmetric equivalence to house-choice
- Equilibrium among utility-maximizing agents who can change jobs or homes through life
- Empirically supported (DePalma, Picard, Waddell 2007 • Bernardin, Trevino, Gliebe 2015 • Gibb 2023)

Clearinghouse models that are NOT EQUIVALENT

- Ordered choice from remaining = *serial dictatorship*
- ActivitySim – similar to *rank maximal*

Better represent society? – Or just a quicker computation?

NON SHADOW-PRICE METHODS
THAT HAVE COME TO SOME ABMS

- Serial dictatorship (seniority, priority)
  - Early Daysim (later changed to shadow-pricing)
  - Original Emme Agent
  - Some college admissions clearinghouses
- Rank maximal (greedy, immediate acceptance)
  - ActivitySim is similar
  - Boston public schools before 2005
- *Used in ABMs for computational reasons*
- *No claims as better representations of how society works.*
- *Distinctly different models with different results.*