Warsaw University
Warsaw Ecological Economics Center

An insight into the numerical simulation bias –

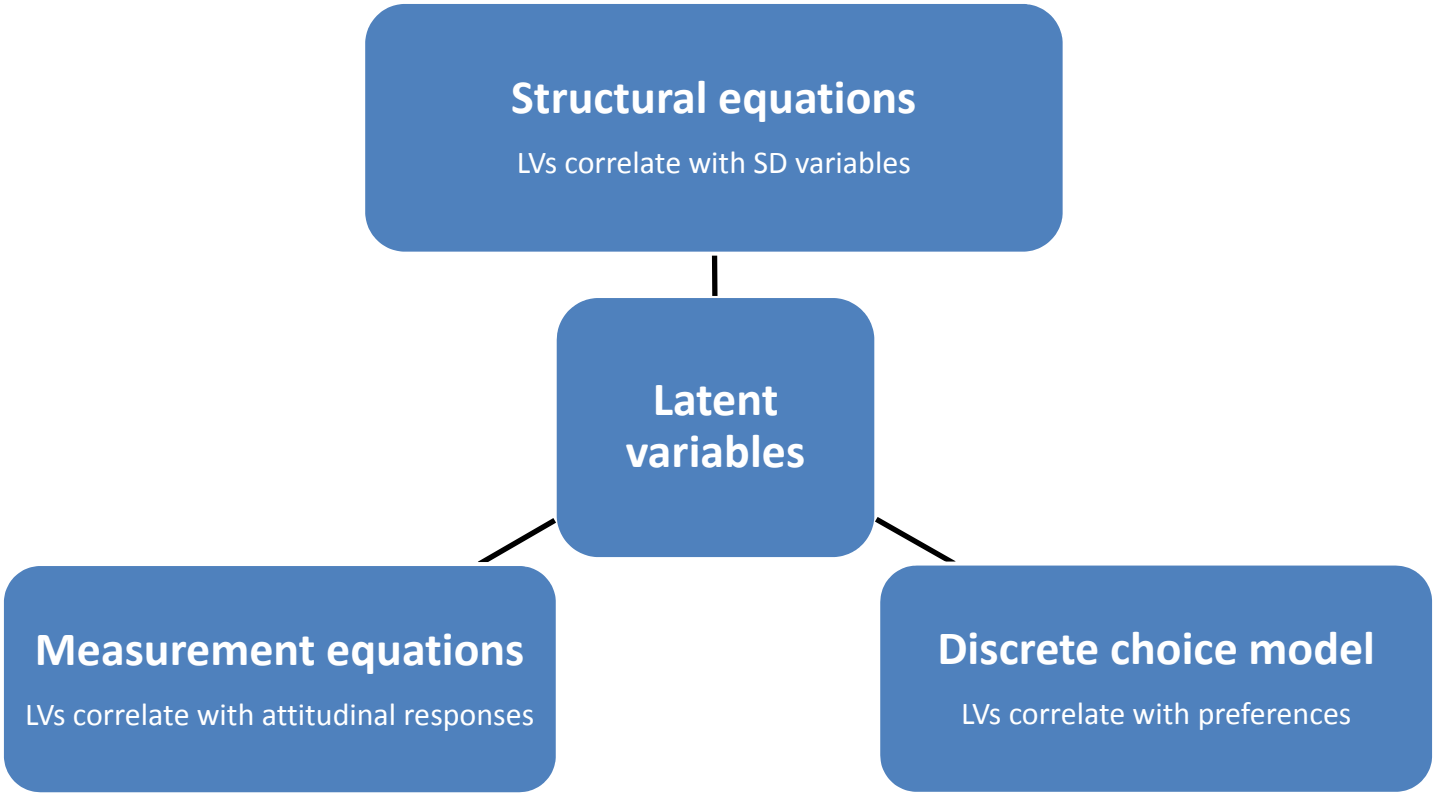
a comparison of efficiency and performance of different types of quasi Monte Carlo simulation methods under a wide range of experimental conditions

Wiktor Budziński
budzinski.wiktor@gmail.com

Mikołaj Czajkowski
miq@wne.uw.edu.pl

Hybrid models – work in progress …

- Should we go hybrid? Investigating the extent of the bias resulting from direct incorporation of attitudinal measures into stated choice models
- The role of objectively defined and subjectively perceived consequentiality for field stated preference studies regarding the provision of public goods



Rationale for studying simulation bias

- Maximum Simulated Likelihood
  - Alternatives: Bayesian estimation (Train and Sonnier, 2005), Expectation-Maximization algorithm (Train, 2007), Laplace approximation (Harding and Hausman, 2007), Maximum Approximate Composite Marginal Likelihood (Bhat, 2011)
  - MSL still the estimator of choice of most choice modelers
- Quasi Monte-Carlo methods
  - Use ‘smarter’ draws to reduce computational burden
  - Train (2000): 100 Halton draws leads to smaller bias and standard deviation of parameter estimates than 1,000 pseudorandom (PMC) draws
  - Bhat (2001): 100 Halton outperform 2,000 PMC draws
  - Later investigations less ‘optimistic’ but the idea is out there and many applied researchers typically use 200-500 Halton draws

Performance of the alternative methods

- Empirical comparisons and alternative methods
  - Halton draws – poor performance in higher dimensions
    - Scrambling or shuffling of the sequence (Bhat 2003, Daly et. al. 2003, Hess and Polak 2003, Wang and Kockelman 2008)
  - Randomized (t,m,s) nets (Sandor and Train 2004)
  - Sobol sequence (Garrido 2003)
  - Modified Latin Hypercube Sampling (Hess et. al. 2006)
  - Randomized lattice rules (Munger et. al. 2012)
  - Generalized antithetic draws with double base shuffling (Sidharthan and Srinivasan 2010)

Problems with existing studies

- Small number of draws
  - If 200 outperforms 1,000 – is 200 enough?
  - Low number of draws can mask identification problems (Chiou and Walker 2007)
  - Even 1,000 antithetic Halton draws can interfere with LR inference (Andersen 2013)
- Small number of repetitions
  - E.g., Bhat (2003), Sandor and Train (2004), Garrido (2003) and Hess et. al. (2006) used no more than 10 repetitions
- A single dataset
  - No. of choice tasks per respondent, no. of respondents
  - No. of attribute levels, no. of attributes, no. of alternatives

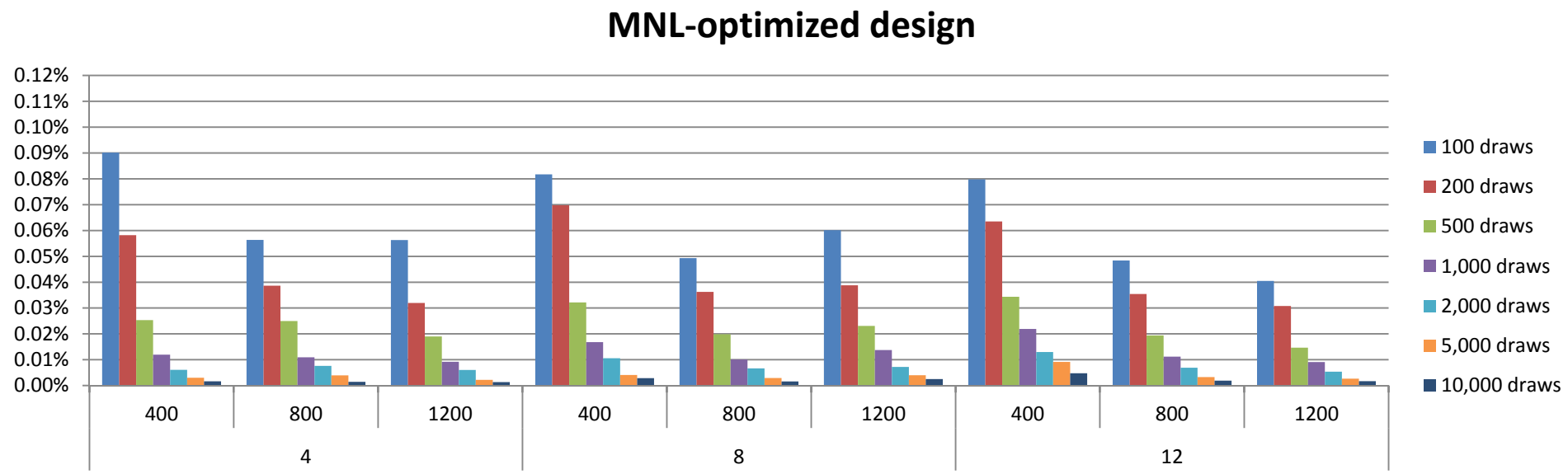
Methodology of our simulation study – estimation

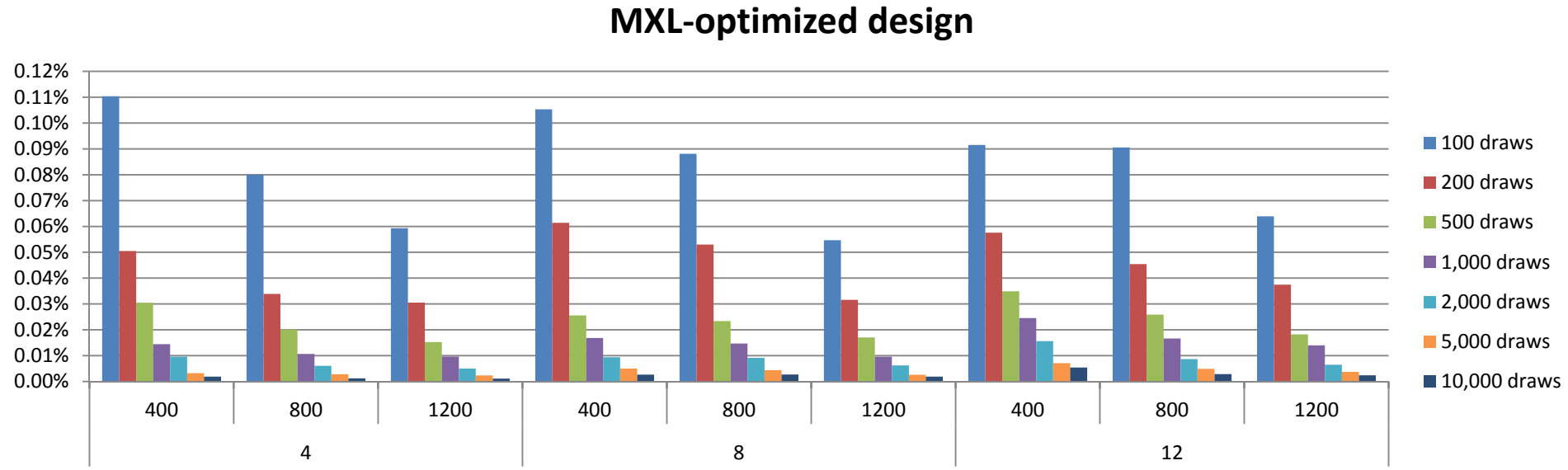
- 5 types of draws:
  - Pseudo Monte Carlo (PMC)
  - Latin Hypercube Sampling (LHS)
  - Randomized scrambled Halton sequence – set-specific shifts (RSH1)
  - Randomized scrambled Halton sequence – respondent-specific shifts (RSH2)
  - Randomized scrambled Sobol sequence (SOB)
- Number of draws
  - 100; 200; 500; 1,000; 2,000; 5,000; 10,000
  - + 100,000 for baseline (SOB only)
- 100 repetitions of each model

Methodology of our simulation study – design

- Number of choice tasks
  - 4; 8; 12
- Number of respondents
  - 400; 800; 1,200
- MXL model
  - 5 normally distributed parameters
    - 1 continuous (4 levels + 0), 3 binary, 1 ASC
  - No correlations
- 2 types of design
  - D-efficient, optimized for MNL or MXL
  - 3 alternatives (SQ + 2 improvements)
  - 2 blocks
- 100x5x7x3x3x2 = 63,000 estimated models

How sure are we that the LL we get is correct? (standard deviation of LL / mean LL – Sobol only)





How sure are we that the LL we get is correct?

Minimum number of draws for the 95% certainty that P(observed LL = true LL) – for different significance levels of LR test (d.f. = 1)

			1%		5%		10%	
Design	CT	NP	No. of draws	Type	No. of draws	Type	No. of draws	Type
MNL	4	400	100	LHS, RSH1, RSH2, SOB	100	LHS	200	LHS,SOB
MNL	4	800	100	SOB	200	SOB	500	RSH2,SOB
MNL	4	1200	200	RSH2,SOB	500	RSH2,SOB	500	SOB
MNL	8	400	200	LHS, RSH1, RSH2, SOB	500	RSH1,RSH2,SOB	1000	RSH1,RSH2,SOB
MNL	8	800	200	SOB	500	SOB	1000	RSH2,SOB
MNL	8	1200	500	SOB	1000	SOB	2000	RSH2,SOB
MNL	12	400	500	LHS, RSH1, RSH2, SOB	1000	RSH2,SOB	1000	SOB
MNL	12	800	500	RSH2,SOB	1000	RSH1,SOB	2000	RSH1,RSH2,SOB
MNL	12	1200	500	SOB	1000	SOB	2000	RSH2,SOB
MXL	4	400	200	RSH1,RSH2,SOB	200	SOB	500	RSH1,RSH2,SOB
MXL	4	800	200	SOB	500	RSH2,SOB	500	SOB
MXL	4	1200	200	SOB	500	SOB	1000	RSH2,SOB
MXL	8	400	500	LHS, RSH1, RSH2, SOB	500	SOB	1000	RSH1,SOB
MXL	8	800	500	SOB	1000	SOB	2000	RSH1,SOB
MXL	8	1200	500	SOB	2000	RSH1,RSH2,SOB	2000	SOB
MXL	12	400	1000	RSH2,SOB	2000	RSH1,RSH2,SOB	5000	RSH1,RSH2,SOB
MXL	12	800	1000	SOB	2000	SOB	5000	RSH1,RSH2,SOB
MXL	12	1200	2000	RSH1,RSH2,SOB	2000	SOB	5000	RSH1,RSH2,SOB

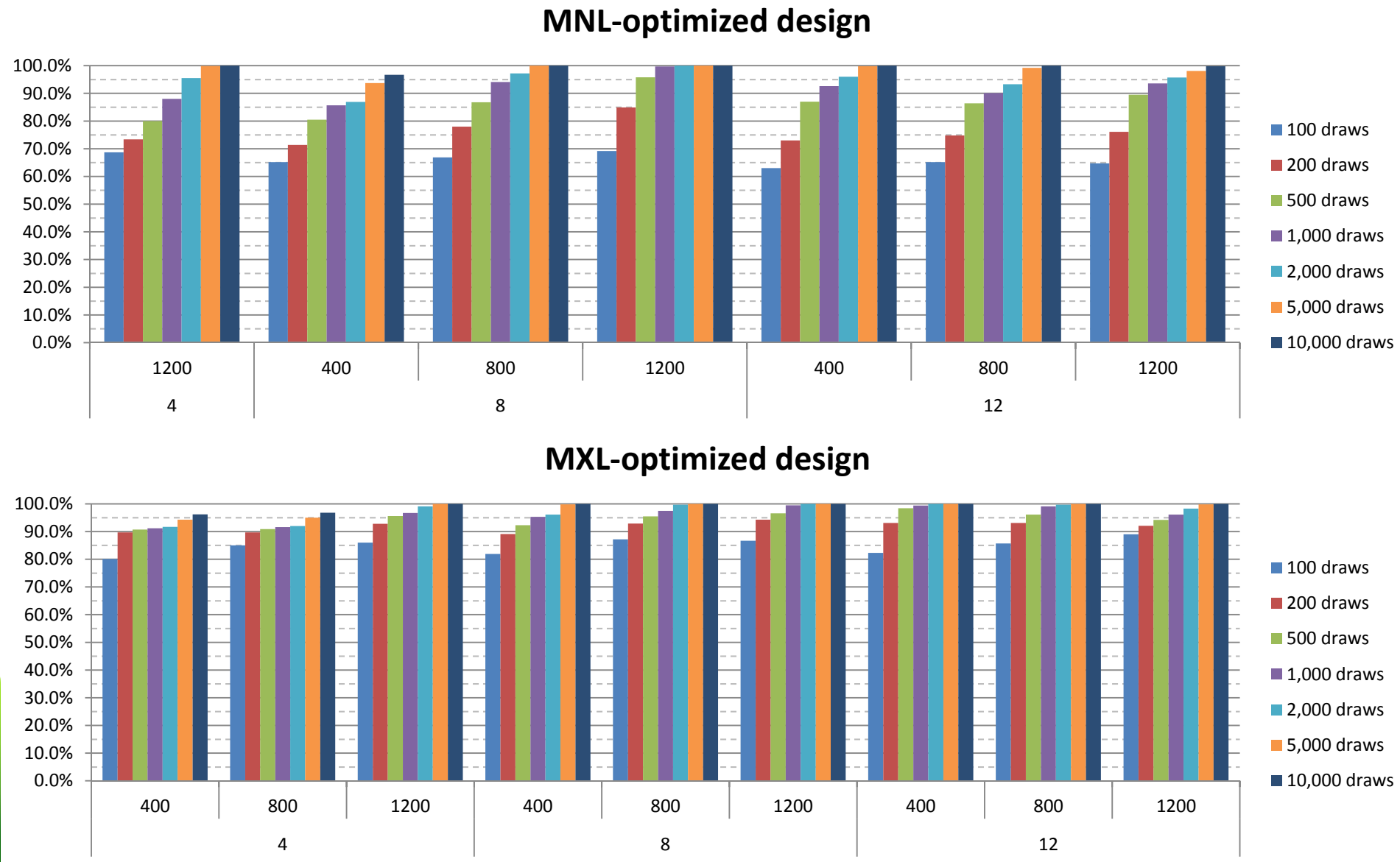
How sure are we that the ***B*** we get are correct?

Minimum number of draws for 95% confidence the Median Absolute Percentage Error (MAPE) less than …

			10%		5%		1%	
Design	CT	NP.	No of draws	Type	No of draws	Type	No. of draws	Type
MNL	4	1200	1000	RSH2,SOB	2000	SOB	>10000	(10,000 SOB = 90%)
MNL	8	400	5000	RSH1,RSH2,SOB	10000	SOB	>10000	(10,000 SOB =84%)
MNL	8	800	1000	RSH1,RSH2,SOB	2000	RSH2,SOB	>10000	(10,000 SOB =93%)
MNL	8	1200	200	RSH2,SOB	500	SOB	5000	SOB
MNL	12	400	1000	RSH1,RSH2,SOB	2000	SOB	>10000	(10,000 SOB =92%)
MNL	12	800	1000	RSH1,RSH2,SOB	5000	LHS, RHS1, RHS2, SOB	>10000	(10,000 SOB =91%)
MNL	12	1200	500	SOB	2000	RSH2,SOB	>10000	(10,000 SOB =94%)
MXL	4	400	5000	RSH1,RSH2,SOB	10000	SOB	>10000	(10,000 SOB =91%)
MXL	4	800	5000	RSH1,SOB	10000	RSH1,SOB	>10000	(10,000 SOB =92%)
MXL	4	1200	200	RSH2	500	SOB	10000	RSH1,RSH2,SOB
MXL	8	400	1000	LHS, RHS1, RHS2, SOB	1000	SOB	10000	SOB
MXL	8	800	200	PMC,LHS, RHS1, SOB	500	SOB	5000	SOB
MXL	8	1200	100	LHS,SOB	500	all types	5000	RSH1,RSH2,SOB
MXL	12	400	200	all types	500	all types	5000	RSH1,RSH2,SOB
MXL	12	800	200	all types	500	RSH1,RSH2,SOB	5000	RSH2,SOB
MXL	12	1200	500	PMC,RHS1, RHS2, SOB	1000	RSH2,SOB	10000	RSH2,SOB

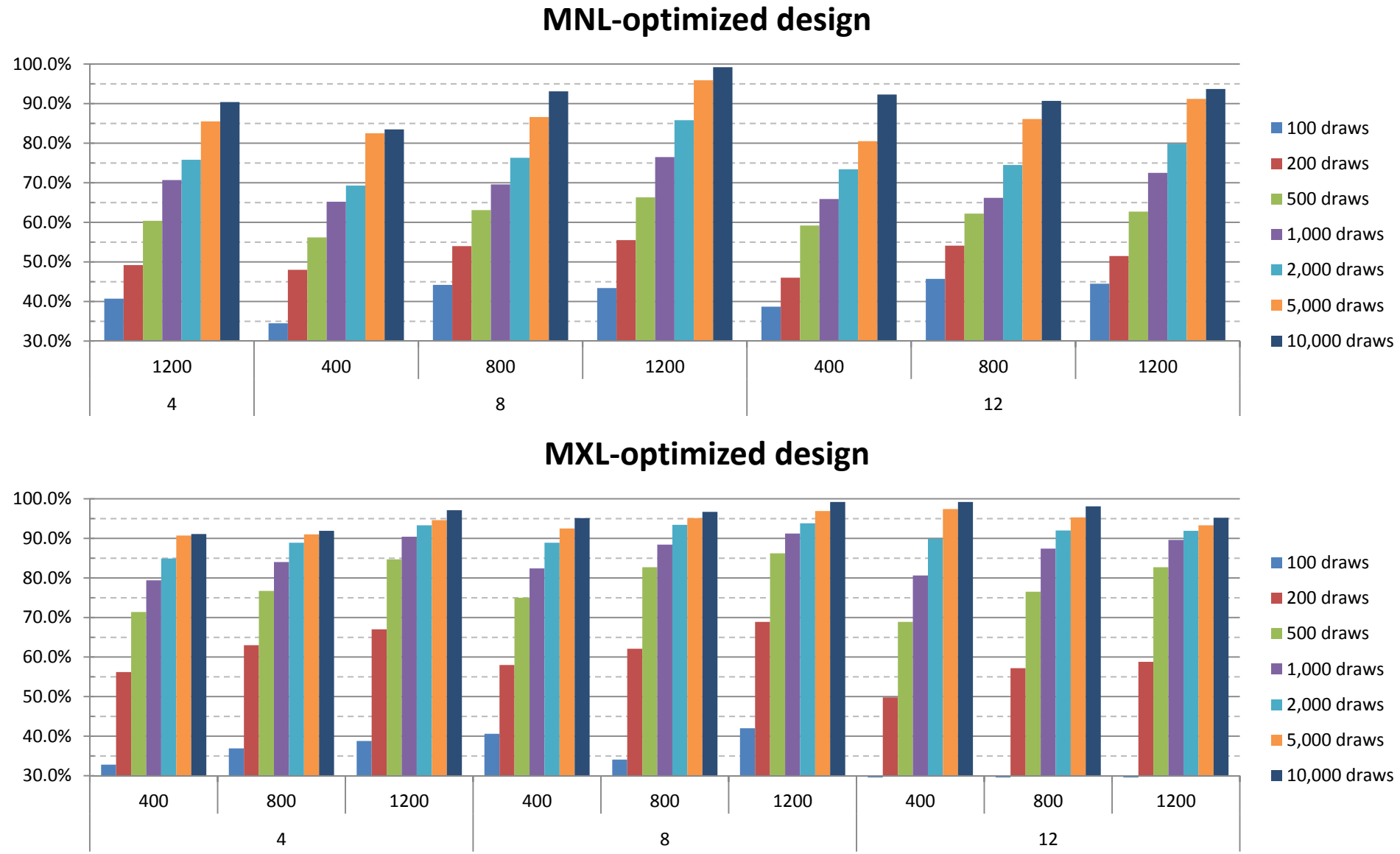
How sure are we that the ***B*** we get are correct?

Probability that MAPE < 5% (Sobol draws only)



How sure are we that the ***B*** we get are correct?

Probability that MAPE < 1% (Sobol draws only)



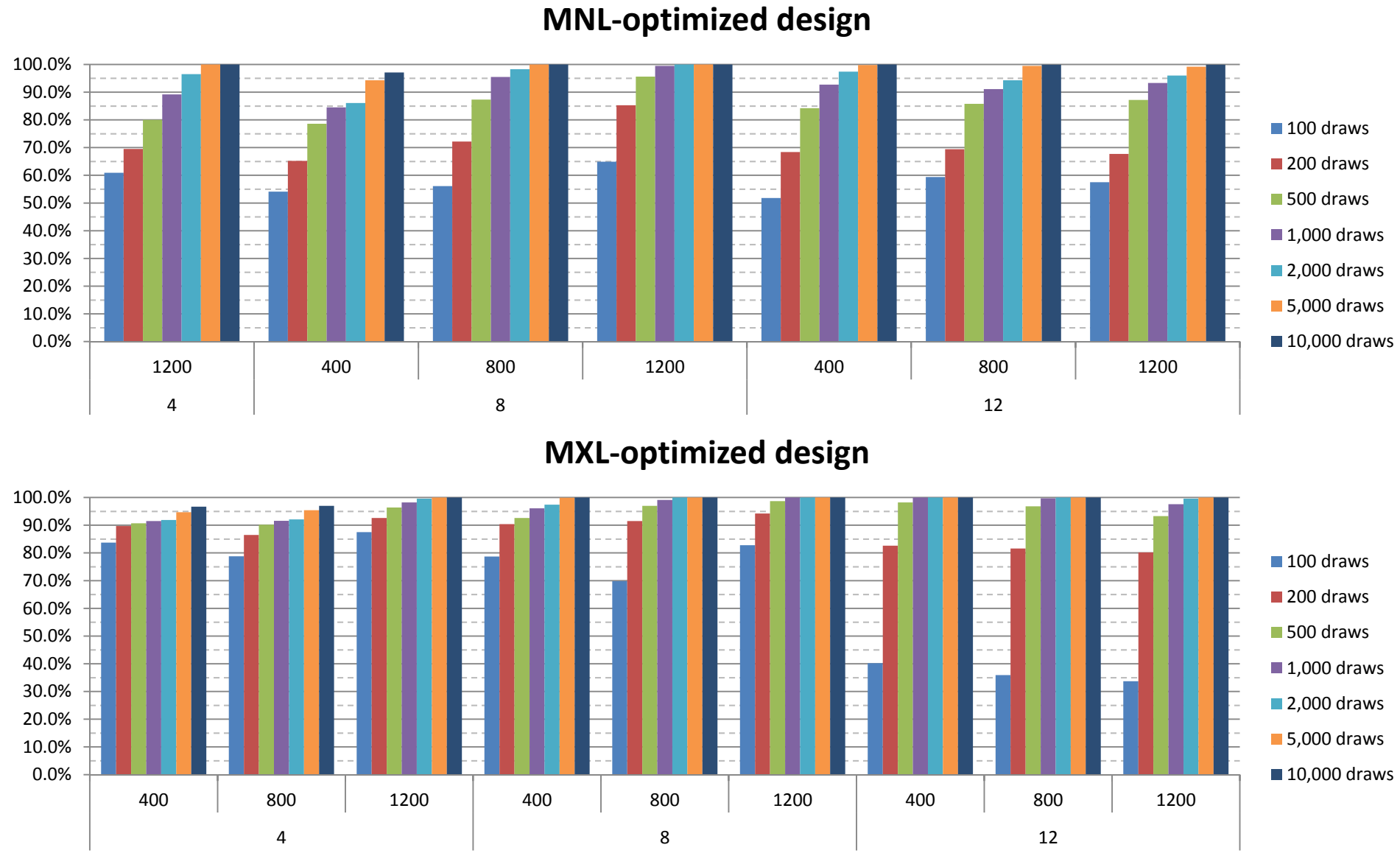
How sure are we that the ***s.e.*** we get are correct?

Minimum number of draws for 95% confidence the Median Absolute Percentage Error (MAPE) less than …

			10%		5%		1%	
Design	CT	NP	No. of draws	Type	No. of draws	Type	No. of draws	Type
MNL	4	1200	1000	RSH2,SOB	2000	SOB	>10000	(10000 SOB = 93%)
MNL	8	400	5000	RSH1,RSH2,SOB	10000	RSH1,RSH2,SOB	>10000	(10000 SOB = 84%)
MNL	8	800	500	SOB	1000	SOB	>10000	(10000 SOB = 94%)
MNL	8	1200	200	LHS, RHS1, RHS2, SOB	500	RSH1,RSH2,SOB	5000	RSH2
MNL	12	400	1000	LHS, RHS1, RHS2, SOB	2000	RSH1,RSH2,SOB	>10000	(10000 SOB = 88%)
MNL	12	800	1000	RSH1,RSH2,SOB	2000	RSH2	>10000	(10000 SOB = 92%)
MNL	12	1200	500	SOB	2000	RSH1,RSH2,SOB	>10000	(10000 SOB = 94%)
MXL	4	400	5000	RSH1,RSH2,SOB	10000	RSH1,SOB	>10000	(10000 SOB = 91%)
MXL	4	800	5000	RSH1,RSH2,SOB	5000	SOB	>10000	(10000 SOB = 92%)
MXL	4	1200	200	RSH2,SOB	500	SOB	5000	RSH2,SOB
MXL	8	400	500	RSH1,SOB	1000	SOB	10000	SOB
MXL	8	800	200	RSH1,RSH2,SOB	500	RSH1,RSH2,SOB	5000	RSH1,RSH2,SOB
MXL	8	1200	200	RSH1,RSH2,SOB	500	LHS, RHS1, RHS2, SOB	2000	SOB
MXL	12	400	200	LHS, RHS1, RHS2, SOB	500	LHS, RHS1, RHS2, SOB	5000	RSH1,RSH2,SOB
MXL	12	800	200	RSH1,RSH2,SOB	500	RSH1,RSH2,SOB	5000	RSH1,RSH2,SOB
MXL	12	1200	500	RSH1,RSH2,SOB	1000	RSH1,RSH2,SOB	10000	RSH1,RSH2,SOB

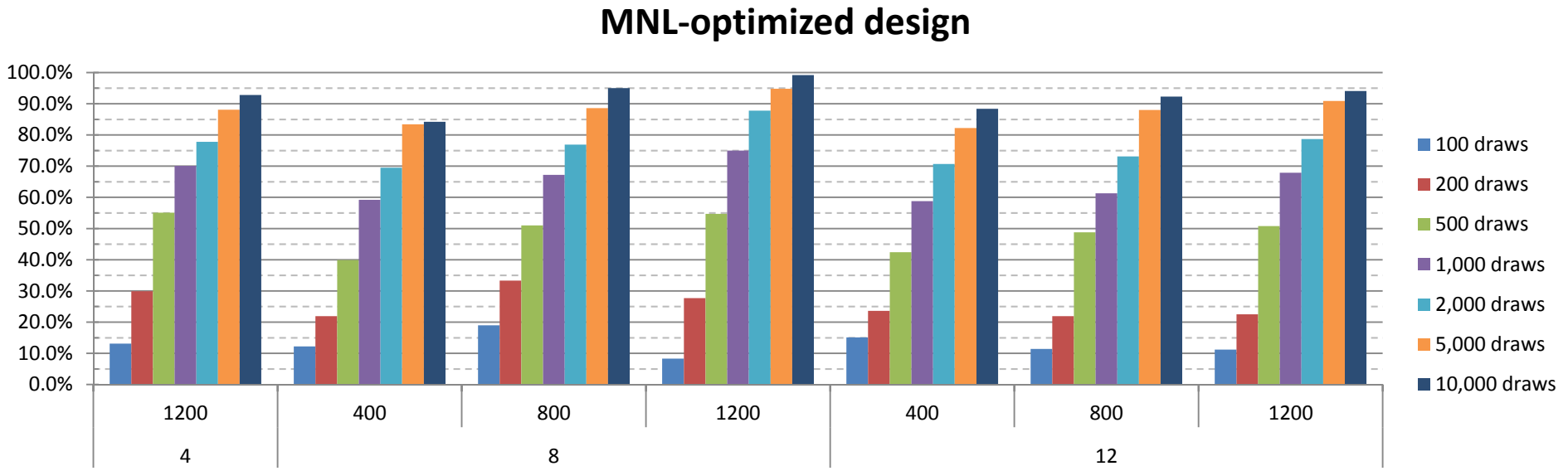
How sure are we that the ***s.e.*** we get are correct?

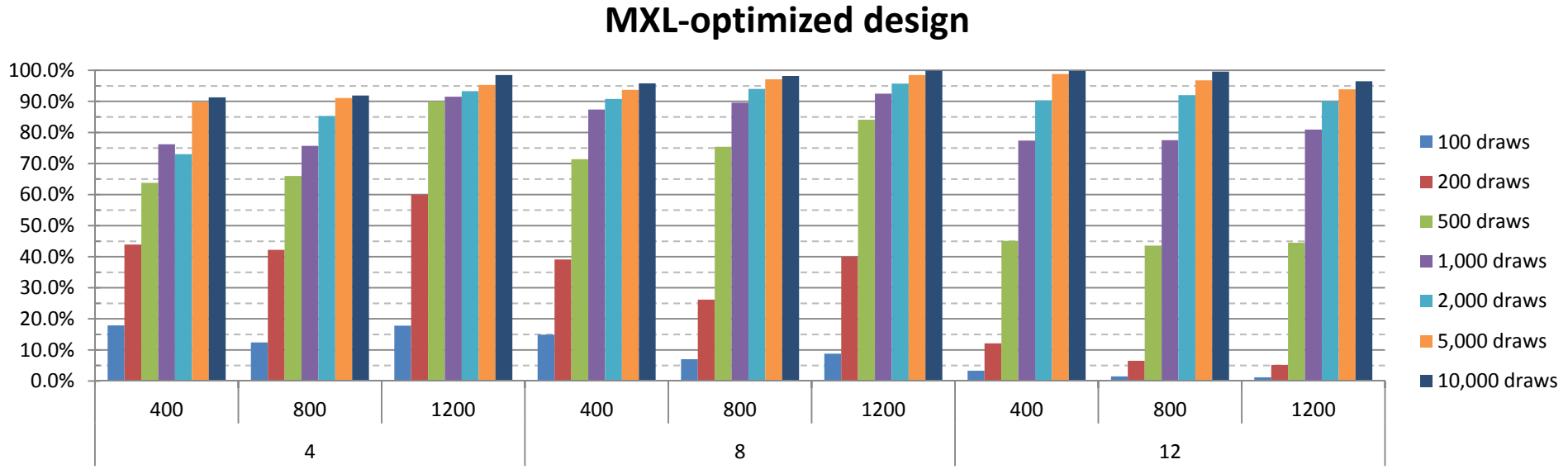
Probability that MAPE < 5% (Sobol draws only)



How sure are we that the ***s.e.*** we get are correct?

Probability that MAPE < 1% (Sobol draws only)





Convergence problems

Mean z-stats (b/s.e.) for 100k Sobol draws models

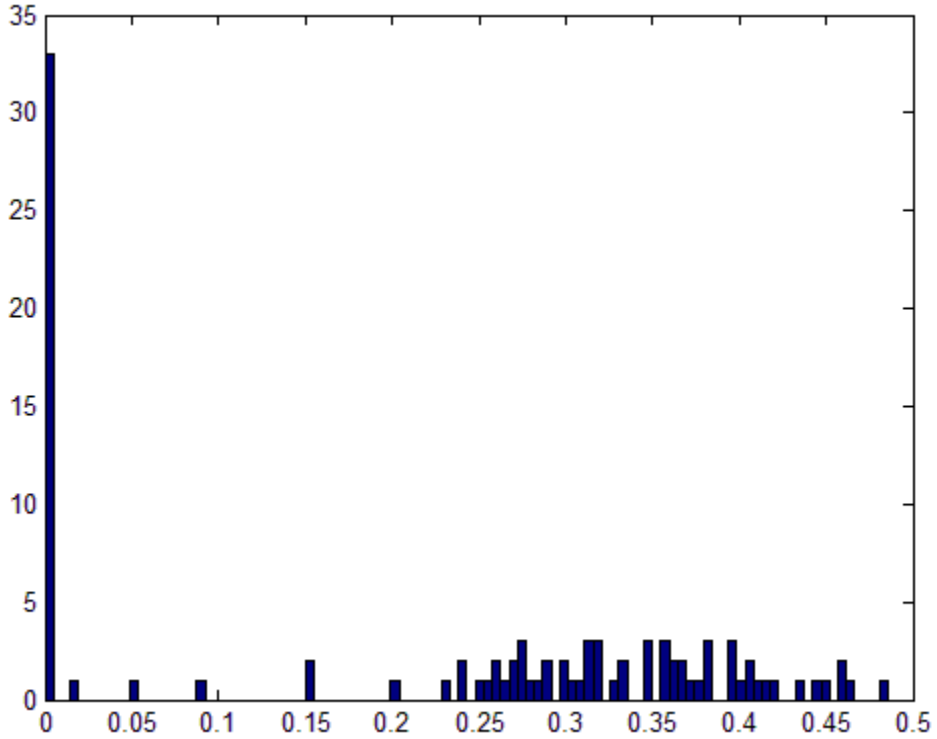
	MNL									MXL								
	4			8			12			4			8			12		
	400	800	1200	400	800	1200	400	800	1200	400	800	1200	400	800	1200	400	800	1200
SQ	6.39	9.01	9.66	9.15	11.95	12.87	10.68	14.36	18.03	5.02	6.98	10.14	10.19	12.37	13.54	9.56	15.56	20.08
ATT1	8.38	11.73	14.21	12.08	16.14	18.44	14.19	19.34	24.54	10.97	16.00	19.37	19.95	24.37	28.99	21.70	32.09	36.74
ATT2	7.98	9.68	12.49	11.04	17.86	18.88	14.48	20.50	23.85	12.88	16.75	22.16	19.24	23.80	27.52	20.16	29.07	35.64
ATT3	8.59	12.44	13.68	11.57	15.02	18.41	15.26	19.47	21.96	11.83	15.56	19.54	18.95	22.69	28.92	20.49	29.95	35.61
FEE	11.61	17.33	20.94	16.51	23.82	30.82	21.02	28.67	35.00	13.87	19.91	23.16	21.71	28.66	35.54	23.04	32.28	40.61
s.d. SQ	3.45	-	1.47	0.81	3.40	5.44	2.67	2.83	2.05	0.55	0.63	2.33	1.64	3.77	4.62	5.91	5.22	3.88
s.d. ATT1	-	1.12	1.43	0.95	2.57	6.69	3.76	5.13	7.08	6.47	6.36	8.59	8.01	12.70	13.23	9.05	13.74	19.01
s.d. ATT2	-	5.19	5.16	4.41	2.07	6.73	5.28	5.11	5.15	5.55	8.78	8.75	7.27	12.22	13.41	10.85	15.94	20.10
s.d. ATT3	0.70	1.92	2.05	3.38	4.51	6.11	2.12	2.07	6.60	4.87	7.88	9.08	6.60	12.93	13.52	9.16	13.75	18.65
s.d. FEE	7.72	12.37	14.20	11.92	16.51	19.23	15.30	19.64	22.82	8.61	11.51	14.70	13.32	17.50	21.94	14.25	20.02	24.10

Convergence problems

1. Models empirically unidentified (Chiou and Walker 2007)
   - Singular Hessians, masked by a low number of draws
2. Models with unrecovered parameters
   - Particularly for estimates of s.d.
   - More draws help uncover the situation

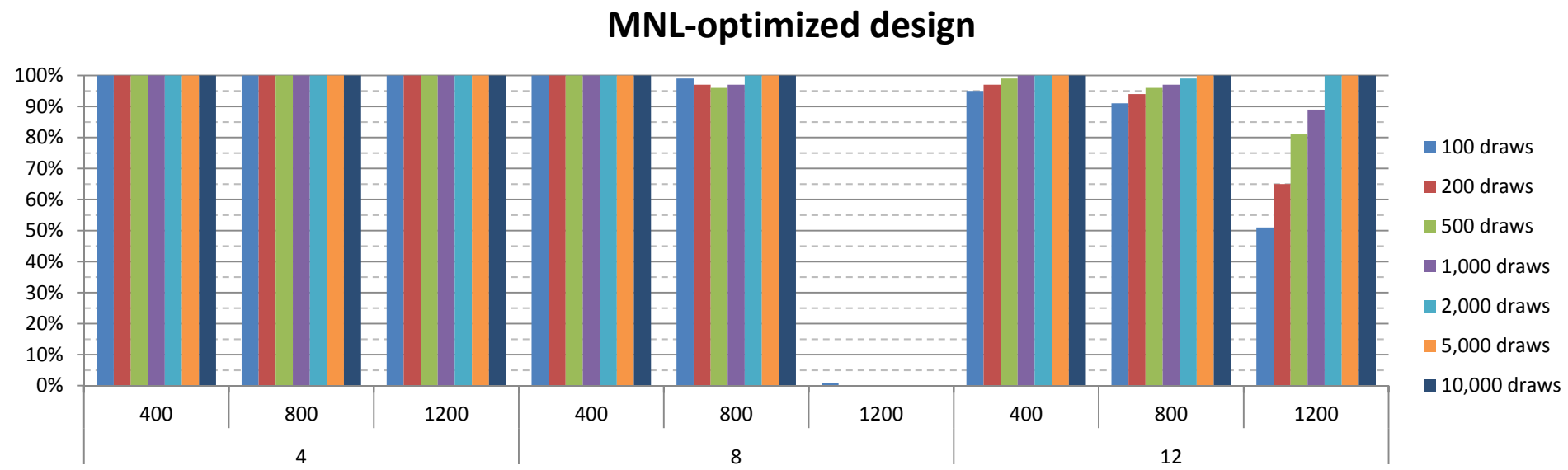
Convergence problems

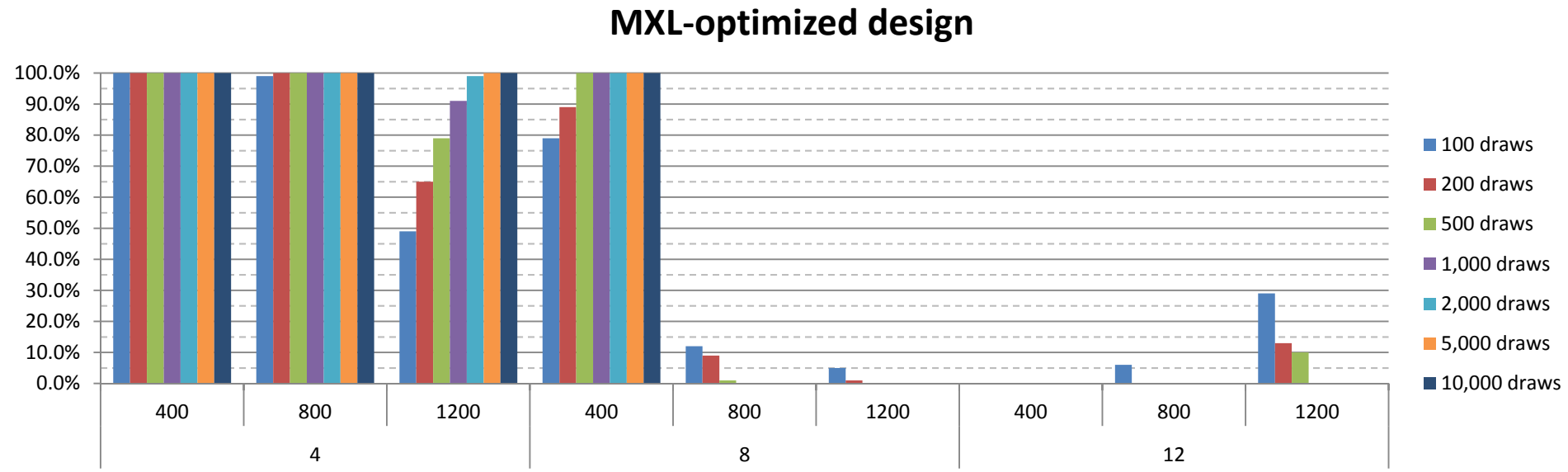
Distribution of the recovered s.d. parameter of SQ (200 Sobol draws, MXL design CT = 4, NP = 800)



Convergence problems

Probability that at least one parameter not significant at the 5% level





tl;dr

- Use scrambled Sobol
- Use at least 2,000 draws (95%)
  - To be on the safe side – use more 10,000?
  - The minimum is design dependent
- More generally
  - Make sure the results are robust w.r.t. increasing the number of draws
  - Watch for identification problems
- Open questions
  - Identification
  - Number of attributes / attribute levels / alternatives / ASCs

czaj.org