

Duke University Synthetic Data Project
Threshold Verification Measure
Simulated Threshold Measure Distribution as a Function of Number of Partitions and Epsilon
Race Disparity Fixed Effects Model using Small to Large Agency, Occupation Subsets
Laplace, Binomial, Beta and Multinomial Results
2/28/2018, T. Balmat

The following tables summarize mean values from the data set used in the 2/18/2018 *Simulated Threshold Measure Distribution* document using the model

$$y = \beta_0 + \beta_{race} + \beta_{age} \times age + \beta_{age^2} \times age^2 + \beta_{ed} \times ed_{years} + \beta_{year} \quad (1)$$

Agency, occupation, sex, and race combinations appear in tables 1 and 2. Values reported in tables 3, 4, and 5 are mean results from ten iterations of fitting model 1.

Table 1: Large to small authentic observation agency, occupation, sex and race subsets chosen for simulation.

Agency		Occupation		Total Observations
VATA	Veterans Health Admin	620	Practical Nurse	66,394
SZ00	Social Security Admin	105	Insurance Admin	42,709
TD03	Federal Aviation Admin	2152	Air Traffic Ctl	21,416
TR93	Internal Revenue Serv	592	Tax Examining	3,963
DJ09	Off of US Attorney	0905	Gen Attorney	3,748
DJ02	FBI	1811	Criminal Investigating	1,225
LF00	Fed Elec Comm	0905	Gen Attorney	32

Table 2: Computed race coefficients using complete (non-partitioned) subsets for indicated agency, occupation, sex, and race.

Agency	Occupation	Sex	Race	$\hat{\beta}_{auth}$	SE	n(Agency-Occ-Sex)	n(Race)	$\hat{\beta}_{synth}$
VATA	620	F	Race(C)	0.02	0.0007	193,144	66,394	0.03
SZ00	105	F	Race(D)	-0.04	0.0010	305,311	42,709	-0.04
TD03	2152	M	Race(C)	-0.05	0.0018	458,114	21,416	-0.06
TR93	592	F	Race(B)	-0.03	0.0023	165,210	3,963	-0.04
DJ09	905	F	Race(C)	-0.01	0.0027	34,154	3,748	-0.01
DJ02	1811	F	Race(B)	-0.04	0.0046	42,077	1,225	-0.06
LF00	905	F	Race(B)	-0.05	0.0329	705	32	-0.06

Column Definitions

M	Number of partition of authentic data
S1	Number of partitions at threshold after fitting model
Serr	Number of partitions for which estimates from the model could not be computed
lbb	Laplace, binomial, beta result
mn	Multinomial result
p1	Posterior proportion S1 partitions
perr	Posterior proportion partitions with no estimates
eps	Epsilon (0.5, 1, or 2)

Table 3: Laplace, binomial, beta mean results from ten iterations

agency	occupation	sex	race	M	S1	Serr	lbb.p1.eps.0.5	lbb.p1.eps.1	lbb.p1.eps.2
VATA	620	F	Race(C)	30.00	0.00	0.00	0.04	0.03	0.01
SZ00	105	F	Race(D)	30.00	29.60	0.00	0.98	0.97	0.98
TD03	2152	M	Race(C)	30.00	27.60	0.00	0.96	0.92	0.92
TR93	592	F	Race(B)	30.00	24.00	0.00	0.79	0.81	0.80
DJ09	905	F	Race(C)	30.00	14.00	0.00	0.50	0.47	0.46
DJ02	1811	F	Race(B)	30.00	22.00	0.00	0.74	0.76	0.74
LF00	905	F	Race(B)	30.00	3.40	23.70	0.54	0.53	0.54

Table 4: Multinomial mean results from ten iterations, partitions at threshold

agency	occupation	sex	race	M	S1	mn.p1.eps.0.5	mn.p1.eps.1	mn.p1.eps.2
VATA	620	F	Race(C)	30.00	0.00	0.03	0.01	0.02
SZ00	105	F	Race(D)	30.00	29.60	0.83	0.87	0.92
TD03	2152	M	Race(C)	30.00	27.60	0.78	0.84	0.86
TR93	592	F	Race(B)	30.00	24.00	0.69	0.77	0.74
DJ09	905	F	Race(C)	30.00	14.00	0.34	0.43	0.44
DJ02	1811	F	Race(B)	30.00	22.00	0.66	0.71	0.70
LF00	905	F	Race(B)	30.00	3.40	0.08	0.10	0.11

Table 5: Multinomial mean results from ten iterations, partitions with non-computable estimates

agency	occupation	sex	race	M	Serr	mn.perr.eps.0.5	mn.perr.eps.1	mn.perr.eps.2
VATA	620	F	Race(C)	30.00	0.00	0.05	0.02	0.01
SZ00	105	F	Race(D)	30.00	0.00	0.06	0.06	0.01
TD03	2152	M	Race(C)	30.00	0.00	0.08	0.05	0.01
TR93	592	F	Race(B)	30.00	0.00	0.05	0.04	0.03
DJ09	905	F	Race(C)	30.00	0.00	0.11	0.03	0.02
DJ02	1811	F	Race(B)	30.00	0.00	0.07	0.04	0.01
LF00	905	F	Race(B)	30.00	23.70	0.76	0.77	0.75