Measuring Inertia: Appendix

E.G. Alberts and I. Gerostathopoulos

May 2022

1 Extra Results

Policy	**Hyper parameter**	**Context A** ($n = 0$–110)	**Context B** ($n = 112$–222)	**Context A** ($n = 224$–244)	**Weighted Average**
Random	n/a	0.33	0.33	0.35	0.34
EXP3FH	n/a	0.5	0.5	0.11	0.47
UCBTN	n/a	**0.68**	**0.97**	*0.0*	**0.72**
e-greedy	$\epsilon = 0.3$	**0.78**	0.44	0.1	0.59
	$\epsilon = 0.4$	0.69	0.48	0.14	0.56
	$\epsilon = 0.5$	0.66	0.45	0.15	0.54
	$\epsilon = 0.6$	0.57	0.43	0.21	0.48
	$\epsilon = 0.7$	0.51	*0.39*	0.27	0.44
	$\epsilon = 0.8$	0.46	*0.39*	0.24	*0.41*
EXP4	$\eta = 0.04$	0.48	0.56	0.19	0.48
	$\eta = 0.10$	0.56	0.56	0.08	0.51
	$\eta = 0.20$	0.62	0.56	0.07	0.54
	$\eta = 0.40$	0.67	0.51	0.07	0.56
	$\eta = 0.60$	0.69	0.54	0.06	0.57
	$\eta = 0.80$	0.7	0.52	0.05	0.57
D-UCB	$\gamma = 0.89$	*0.39*	0.42	0.4	*0.4*
	$\gamma = 0.92$	*0.42*	0.45	0.42	0.43
	$\gamma = 0.97$	0.47	0.55	0.49	0.5
	$\gamma = 0.99$	0.58	0.74	**0.55**	0.64
	$\gamma = 0.995$	0.61	0.84	0.33	0.67
	$\gamma = 0.997$	0.64	**0.91**	*0.03*	**0.68**

Table 1: XXY: Convergence Rates per Context

- The inertia of UCB-Tuned is shown in Context 3. - DUCB provides an tradeoff between committing and overcoming inertia (better UCB-Tuned). - UCB-Tuned is bad for longer periods of time sequentially which can have larger drawbacks than the intermittent inaccuracies of a non-stationary policy. Reward graph for the best and the worst in each context: UCBTN, DUCB99, eGreedy0.7, DUCB89, Random

Future Work: -XXX?

Policy	**Hyper parameter**	**Context A** ($n = 0-20$)	**Context B** ($n = 22-132$)	**Context A** ($n = 134-154$)	**Weighted Average**
Random	n/a	0.34	0.33	0.34	0.33
EXP3FH	n/a	0.37	0.6	0.03	0.47
UCBTN	n/a	0.5	**0.81**	0.25	**0.67**
egreedy	$\epsilon = 0.3$	**0.6**	0.47	0.2	0.44
	$\epsilon = 0.4$	**0.52**	0.45	0.22	0.42
	$\epsilon = 0.5$	0.51	0.41	0.22	0.39
	$\epsilon = 0.6$	0.48	0.38	0.24	0.37
	$\epsilon = 0.7$	0.47	0.36	0.25	0.36
	$\epsilon = 0.8$	0.46	0.34	0.27	0.35
EXP4	$\eta = 0.04$	0.41	0.67	0.12	0.54
	$\eta = 0.10$	0.44	0.77	0.02	0.59
	$\eta = 0.20$	0.4	0.8	0.02	0.61
	$\eta = 0.40$	0.44	0.82	0.03	0.62
	$\eta = 0.60$	0.46	0.81	0.01	0.62
	$\eta = 0.80$	0.5	0.84	0.03	0.65
DUCB	$\gamma = 0.89$	0.4	0.43	0.4	0.42
	$\gamma = 0.92$	0.4	0.45	0.43	0.44
	$\gamma = 0.97$	0.45	0.56	**0.52**	0.53
	$\gamma = 0.99$	0.49	0.68	**0.55**	0.63
	$\gamma = 0.995$	0.5	0.73	0.43	**0.65**
	$\gamma = 0.997$	0.5	**0.76**	0.35	**0.65**

Table 2: YXY

Policy	**Hyper parameter**	**Context A** ($n = 0-110$)	**Context B** ($n = 112-132$)	**Context A** ($n = 134-154$)	**Weighted Average**
Random	n/a	0.33	0.33	0.3	0.32
EXP3FH	n/a	0.54	0.22	0.62	0.52
UCBTN	n/a	0.69	0.80	0.31	0.64
egreedy	$\epsilon = 0.3$	**0.75**	0.09	**0.81**	**0.7**
	$\epsilon = 0.4$	**0.72**	0.11	**0.72**	**0.67**
	$\epsilon = 0.5$	0.64	0.18	0.68	0.6
	$\epsilon = 0.6$	0.59	0.17	0.57	0.54
	$\epsilon = 0.7$	0.52	0.24	0.53	0.5
	$\epsilon = 0.8$	0.45	0.27	0.46	0.43
EXP4	$\eta = 0.04$	0.47	0.43	0.5	0.47
	$\eta = 0.10$	0.55	0.33	0.6	0.53
	$\eta = 0.20$	0.63	0.25	0.66	0.6
	$\eta = 0.40$	0.65	0.24	0.63	0.61
	$\eta = 0.60$	0.67	0.21	0.73	0.64
	$\eta = 0.80$	0.7	0.21	0.67	0.65
DUCB	$\gamma = 0.89$	*0.39*	0.38	0.41	0.39
	$\gamma = 0.92$	*0.42*	0.46	0.42	0.42
	$\gamma = 0.97$	0.48	0.65	0.48	0.5
	$\gamma = 0.99$	0.59	0.77	0.31	0.56
	$\gamma = 0.995$	0.64	**0.85**	0.25	0.6
	$\gamma = 0.997$	0.65	**0.85**	0.29	0.62

Table 3: XYY

- UCBTuned is stressed more because it doesn't have time to converge in contexts 2 and 3. However, we see the (accidental but positive) effect of inertia in context 3. - Discount UCB quite literally benefits from a direct reduction in inertia by forgetting some values from context 1 which UCB-Tuned cannot.

- There seems to be a sweet spot for the tuning of DUCB