Stability of Measure Comparisons for wig+100 on msmarco-passage (p=0.75) Count Delta 0.5 600 1200 1800 2400 0.4 3000 • Swap Probability 0.1 0.0 0.2 0.6 8.0 1.0 1.2 1.4 0.4 Average Performance Delta