The trend of (E1 - E2) across different pretraining data sizes.