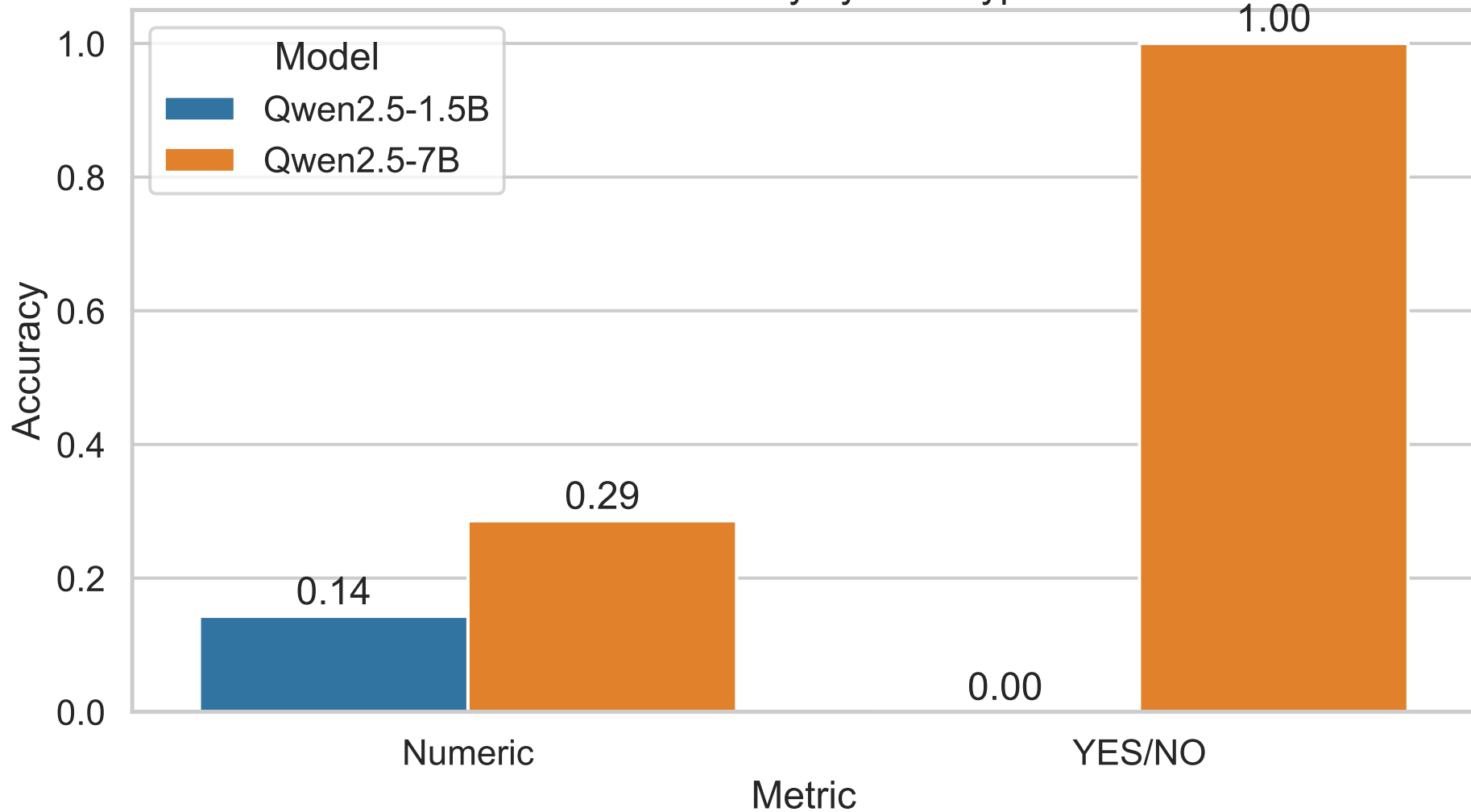
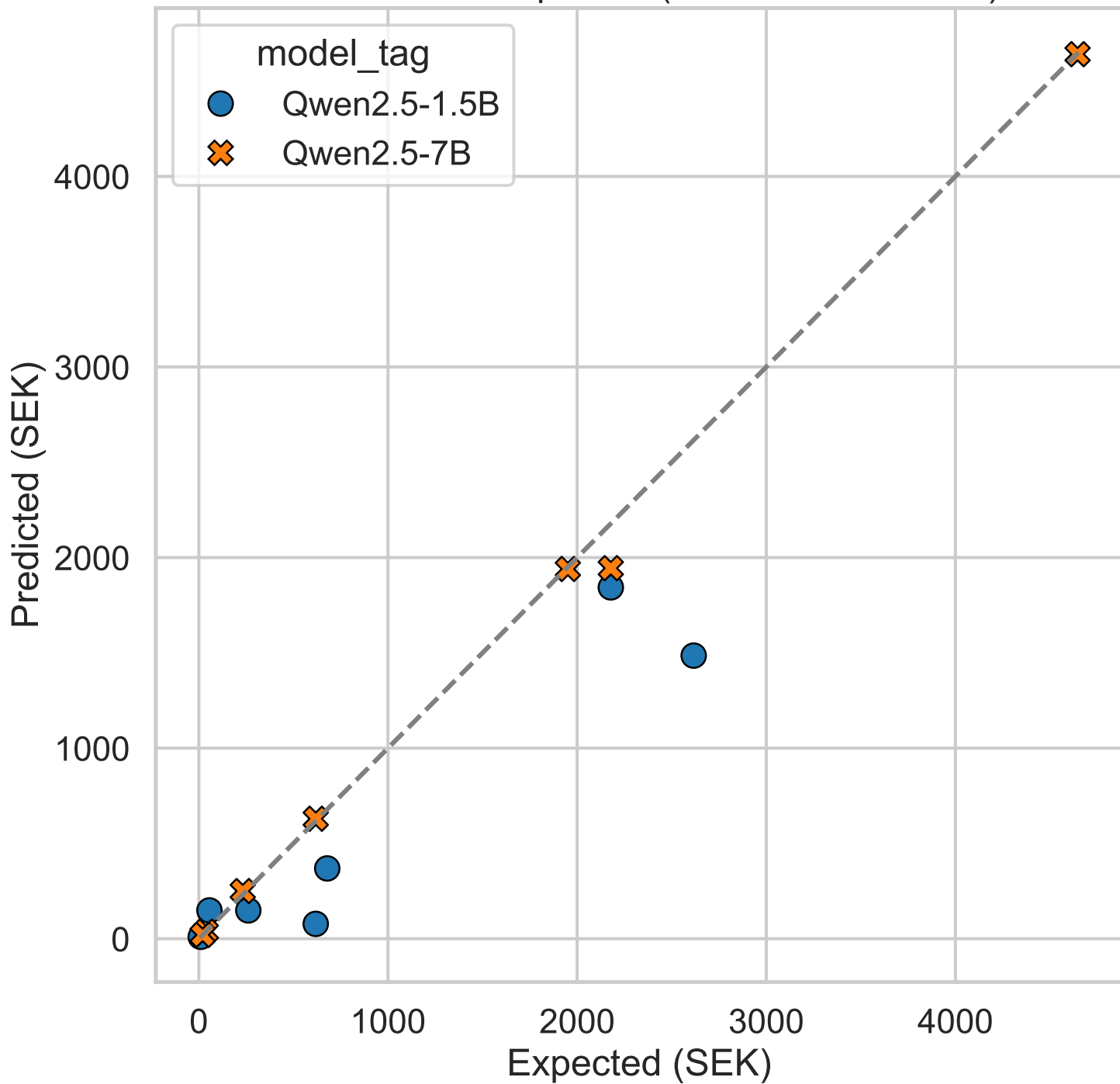


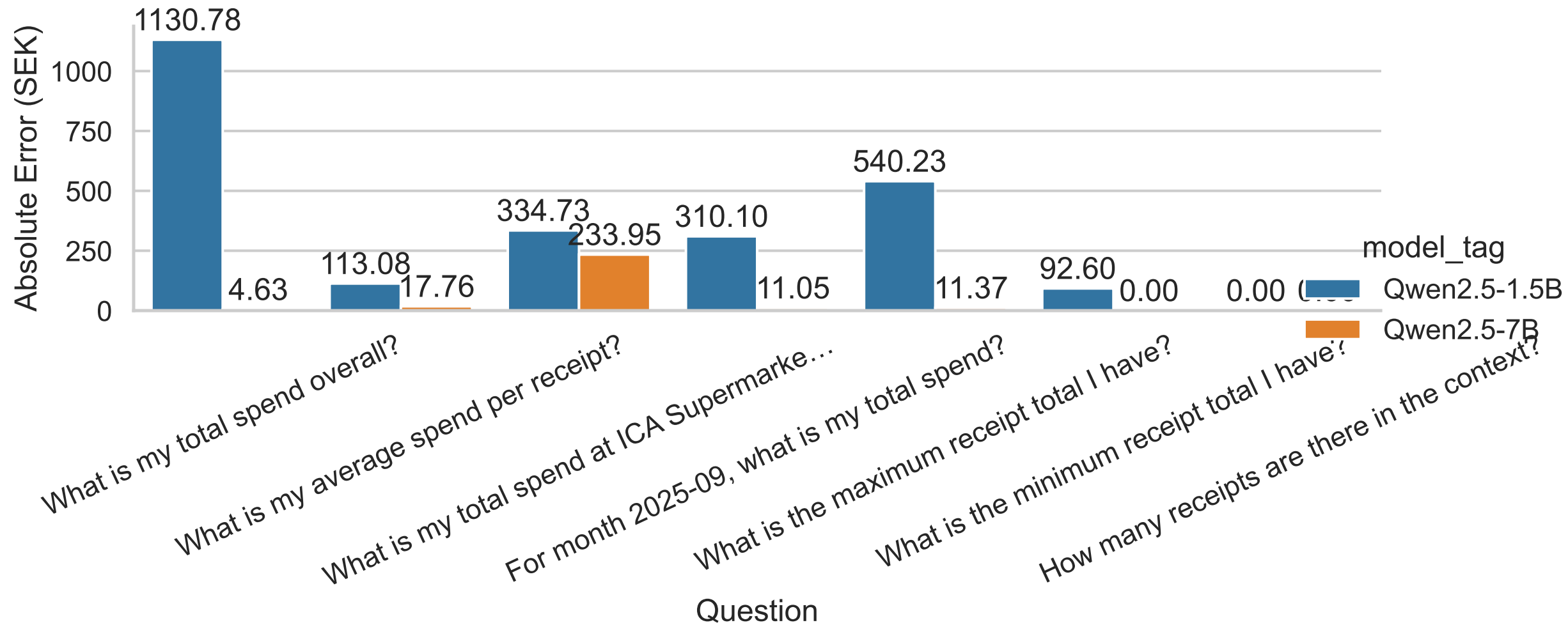
LLM Accuracy by Task Type



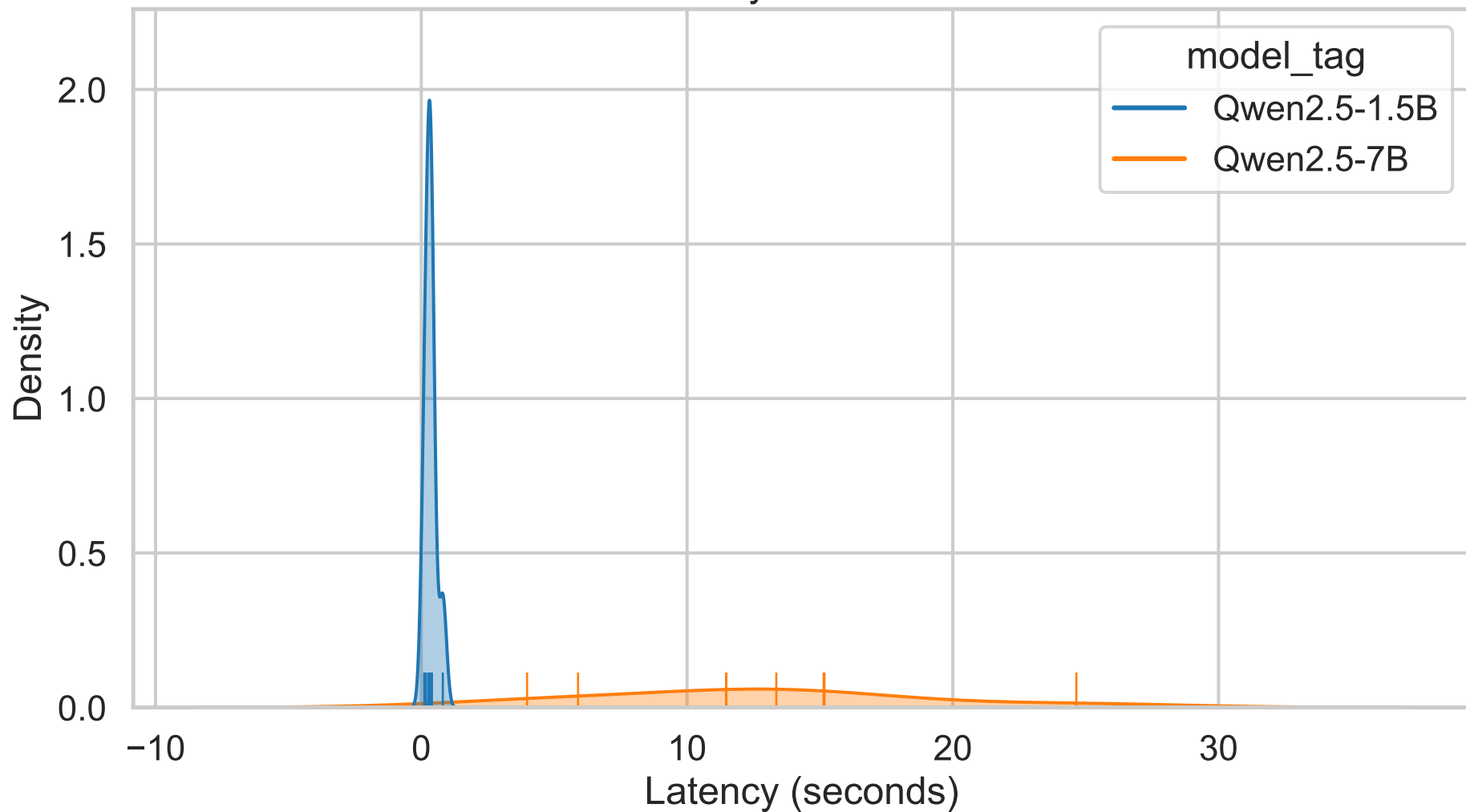
Predicted vs Expected (Numeric Questions)



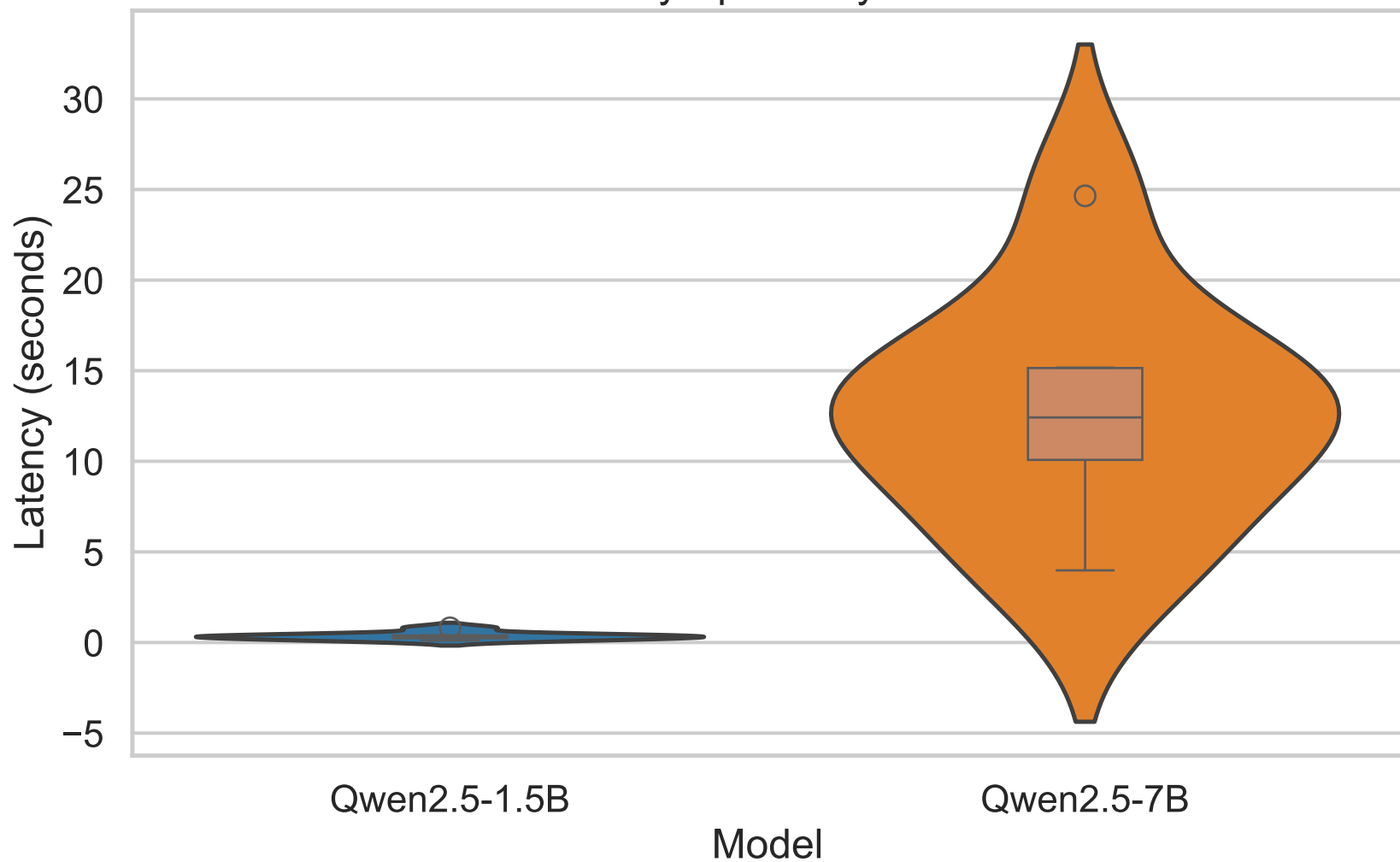
Absolute Error per Numeric Question



Latency Distribution



Latency Spread by Model



# Summary Metrics

Model	Numeric Accuracy	Bool Accuracy	MAE (SEK)	RMSE (SEK)	Within $\pm 1$ SEK	Median Abs Err (SE)	Latency mean (s)	Latency p50 (s)	Latency p95 (s)
Qwen2.5-1.5B	0.143	0.000	360.217	507.104	0.143	310.100	0.352	0.345	0.666
Qwen2.5-7B	0.286	1.000	39.823	88.899	0.286	11.050	12.644	12.420	21.332