Simulator Performance

QB's execution simulator is an important tool for developing and evaluating our algorithms. Starting in November 2013, we have been sending into the simulator a copy of each client order, in supported futures products on the CME, Eurex, and LIFFE (ICE Europe) exchanges. The simulated orders are running exactly the same version of the algorithmic execution code, in identical market conditions, on the same proximity-hosted servers. This thus represents an ideal opportunity to validate the assumptions that underly the simulator, and to compare the relative accuracy of the simulator across different product categories.

The simulator takes account of special features of interest rates futures markets. These markets are generally "large tick:" the actual bid-ask spread is nearly always equal to the minimum price increment imposed by the exchange, and the inside quote sizes are very large. Thus we expect simulator performance to be more accurate for interest rate products than for equity index and energy. The simulator does not include an estimation of market impact and hence is most accurate for orders that are not a large fraction of market volume.

This study reports on the period from January through October 2016. For these 10 months, this data set consists of 90,348 separate parent orders. Individual orders range in size from a single lot to 20,000 lots.

We exclude orders for which the executed quantity in the simulator was different from the production quantity, most frequently because the production order was cancelled before completion or due to client constraints.

We include only orders using QB's Bolt algorithm, whose benchmark is the "arrival price:" the bid-ask midpoint at the time that the order was received. For a buy order, the slippage is the average price per lot achieved for the order minus this benchmark price, with signs reversed for a sell order. Positive slippage is bad, negative is good.

Figure 1 shows slippage for simulator executions versus real slippage, for CME interest rates futures. The scale on each axis is measured in units of the minimum price increment for each product. If the simulator were a perfect reproduction of real trading, then all the points would lie on the diagonal line.

Figure 2 shows the distribution of the simulator error, defined as the slippage on the simulated order minus the slippage on the real order executed in production (the deviation from the diagonal line in Figure 1). For interest rate futures (upper panel), the concentratation about zero indicates that the large majority of simulator results are within 1/4 of a minimum price increment of the real production result. For equity index and energy futures, the simulator is very accurate for the E-mini S&P 500 futures, which is generally single-tick. Its error is larger for the other products, which are generally multi-tick. For these products, QB algorithms make extensive use of intelligent spread improvement, which is not accurately modeled in the simulator. Figure 3 shows cumulative distributions of the same data, which eliminates the need to introduce a bin size and which lets us verify the absence of bias.

Table 1 shows statistics for the collection of orders. If Σ^{prod} and Σ^{sim} are the production and simulator slippage for a given parent order, then $\Delta = \Sigma^{\text{sim}} - \Sigma^{\text{prod}}$ is the error arising from use of the simulator; $\Delta > 0$ means the simulator is pessimistic compared to production. In Table 1, $\bar{\Delta}$ and σ_Δ are the sample mean and standard deviation over the collection of parent orders. They are computed using weighted averages, as though the separate fills within each parent order were independent. For an individual parent order, the expected value of the slippage difference will be this $\bar{\Delta}$, but the standard deviation will be smaller because of diversification across the individual lots. On a parent order of s lots, the standard deviation of slippage difference would be $\sigma_\Delta/\sqrt{s}$ if each lot were an independent fill. The actual standard deviation will be much larger since fills come in highly correlated groups.

The mean slippage differences are generally less than one minimum price increment and have both signs. This suggests that the simulated slippage results are reasonably close to the

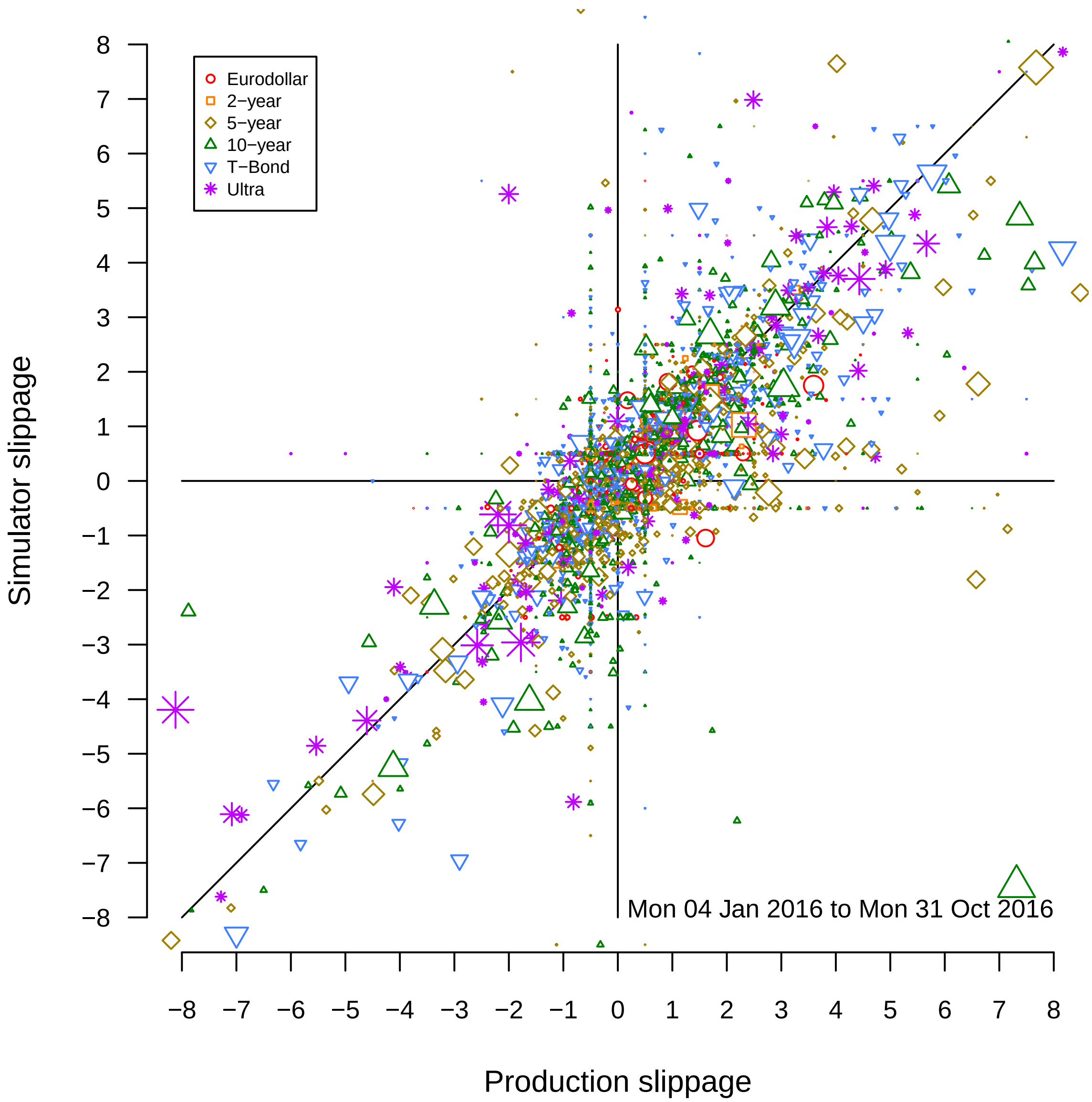

Figure 1: Slippage for orders executed in the simulator, against slippage for identical orders executed at the same time in the real production markets, for interest rates futures contracts. Slippage is measured relative to "arrival price:" the bid-ask midpoint at the time the order was received (negative is good). Axis units are the minimum price increment for each contract; since in these products the bid-ask spread is usually equal to the minimum price increment, $\pm 1/2$ would correspond to execution at the bid or the ask. Symbol size is proportional to parent order size. The analogous plot for equity index and energy products would have wider axis scale to accomodate the larger value of σ_Σ shown in Table 1.

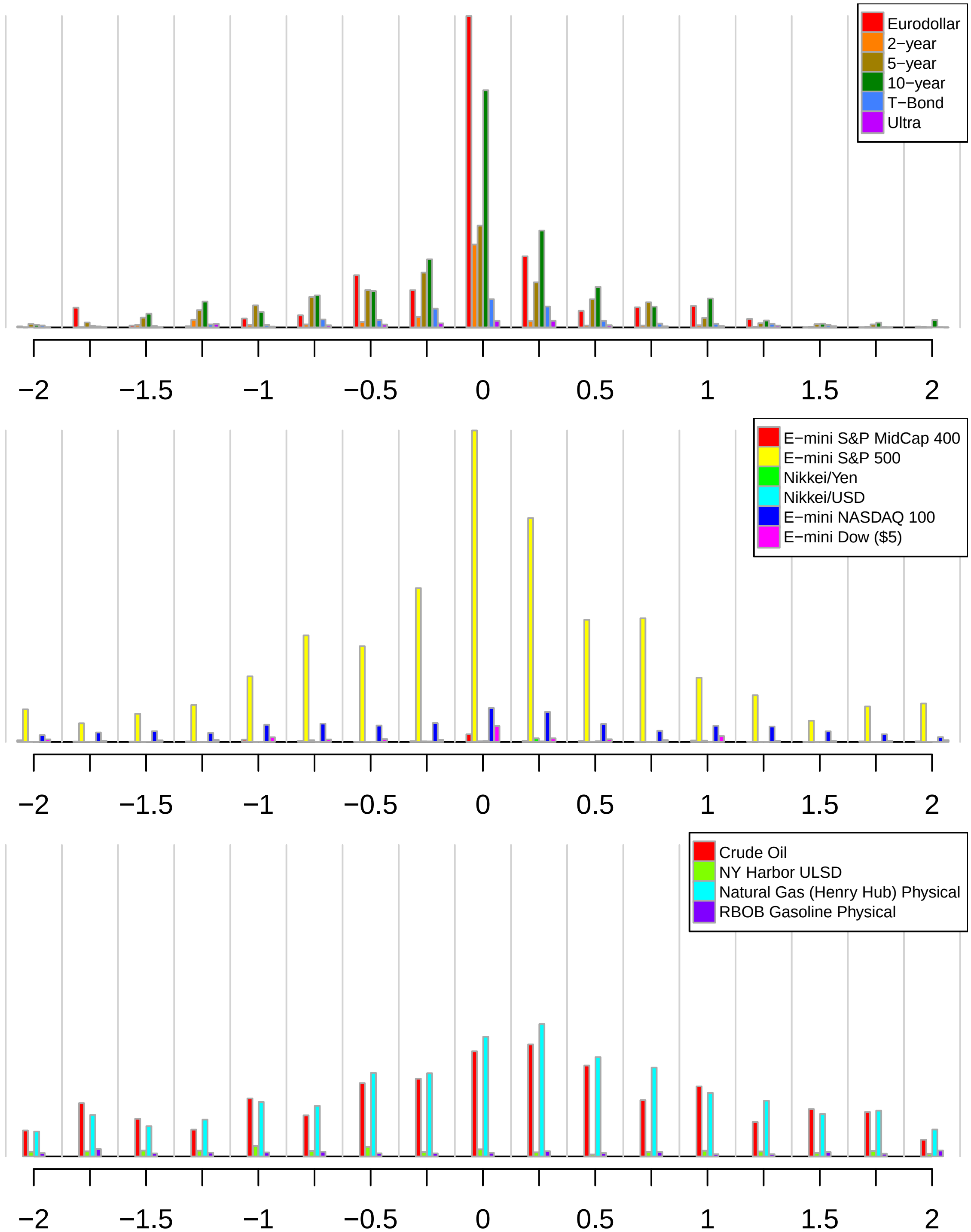

Figure 2: Distribution of difference between real slippage and simulator slippage, in units of minimum price increment. Vertical scale is total number of lots executed in each contract, across all maturities; scale is consistent between different products in each graph, but not between graphs.

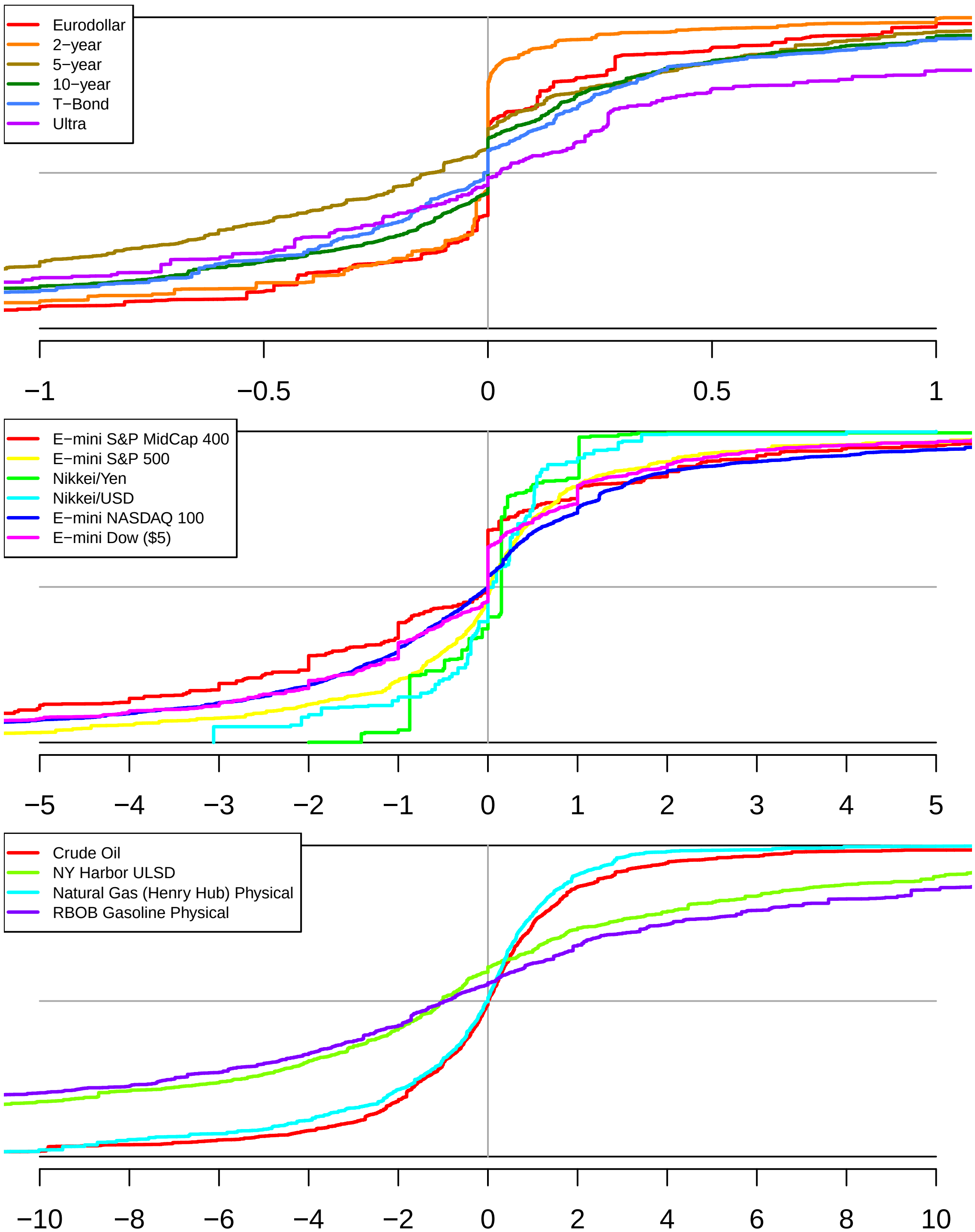

Figure 3: Cumulative distributions corresponding to histograms in Figure 2, with horizontal scale adjusted for each category. The general lack of bias is evident from the passage of the curves through $(0, 0.5)$, except for products with small numbers of orders and a high level of random noise.

Futures Name	px incr ($)	N	$\bar{\Delta}$	σ_Δ	σ_Σ	$\sigma_\Delta/\sigma_\Sigma$
Eurodollar	12.5	3,588	−0.079	0.64	0.89	0.714
2 Year Note	15.625	1,285	−0.151	0.47	0.84	0.558
5 Year Note	7.8125	5,683	−0.560	1.81	2.68	0.675
10 Year Note	15.625	7,645	−0.308	2.03	2.09	0.972
Treasury Bond	31.25	3,192	−0.186	1.12	2.85	0.392
Ultra Bond	31.25	799	0.130	1.53	3.72	0.411
E-mini S&P 500	12.5	9,455	−0.061	3.91	4.05	0.963
E-mini NASDAQ 100	5	8,359	−0.208	5.21	10.75	0.485
E-mini Dow ($5)	5	6,309	−0.431	4.16	5.87	0.709
Crude Oil	10	4,075	−0.304	4.54	5.50	0.826
Natural Gas (Henry Hub)	10	4,027	−0.856	4.12	6.38	0.646
NY Harbor ULSD Heating Oil	4.2	1,962	−2.326	16.75	37.24	0.450
RBOB Gasoline	4.2	1,318	−2.133	24.39	36.79	0.663

Table 1: CME futures products traded and summary statistics. "px incr" is the minimum price increment in dollars, our unit for slippage measurement. N is the number of distinct parent orders in the data sample. $\bar{\Delta}$ is the weighted mean difference between simulator slippage and production slippage, and σ_Δ is its standard deviation. The mean slippage difference is small and unbiased.

actual slippage results, and are unbiased. We do not report a t-statistic for the hypothesis that the mean slippage difference is different from zero, since we do not know how to characterize the standard deviation. The mean slippage differences are generally closer to zero for the interest rates products than for equity index and energy, suggesting that our assumptions are more valid for these products. This is consistent with the focus of the simulator on large-tick products with special emphasis on interest rates.

In Table 1 we also show σ_Σ, the standard deviation of the production slippage itself rather than of the slippage difference, and the ratio of the standard deviation of the slippage difference to the standard deviation of the slippage. These ratios are consistently less than 1, and are not conspicuously smaller for interest rates than for the other products. This suggests that a reason for the large discrepancies in Nasdaq and Dow indices and energy is that these products inherently move more during the time of execution, and often have bid-offer spreads larger than the minimum price increment.

We do not show the mean of the production slippage, although it is the quantity of most economic relevance since it determines the total transaction cost and hence the effectiveness of the algorithms themselves. We do not want to distract attention from our primary focus which is the comparison of the simulator against production.