Understanding Different Sources of Market Impact & Modeling Trading Cost

In this note we present the structure and properties of the trading cost models used by DB Global Equities

July 9, 2018

Introduction

Trading cost models are widely used for pre-trade strategy selection and post-trade performance measurement. They try to explain the cost of consuming liquidity in terms of instrument and order specific variables. Unlike short term microstructure signals, trading cost models work at larger time scales (minutes to hours) and require large order samples for their calibration.

Parametric Cost Models

Trading cost models can be non-parametric or parametric. Non-parametric models are essentially look-up tables where average execution cost is bucketed in bins of the explanatory variables. Parametric models postulate a functional form for the mean expected cost, which depends on the explanatory variables and a few parameters to be estimated statistically.

$$C = f(x;\theta) + \epsilon. \tag{1}$$

- Here, C is the cost versus arrival price, $\bar{C} = f(x;\theta)$ is the mean cost conditional on the explanatory variables x and the model parameters θ, and ϵ is the unexplained random variation.
- The parameters θ are slowly varying with time. In practice they are updated monthly or quarterly. Estimating the parameters is typically done by non-linear regression, which can also be corrected for the heteroscedasticity of ϵ.

Parametric cost models as in eq. (1) are popular because of their simplicity and parsimony. They can be thought of as effective theories summarizing the underlying dynamics of price impact. Their drawback is large unexplained variability.

Model Varieties

DB cost models use instrument and order explanatory variables from the following list:

C	cost vs arrival (in basis points)
s	mean bid-ask spread (in basis points)
σ	daily volatility (in basis points)
V	ADV in shares
Q	order size (shares)
T	order duration (fraction of trading day)
r	order participation rate (fraction of order's filled quantity over market's volume within interval T)

For an order that executes at fixed participation rate (POV-like) and a market with slowly varying volume profile the following relation holds

$$\frac{Q}{V} = rT. \tag{2}$$

For orders with variable participation rate, the right hand side above becomes an integral $\int_0^T r(t)dt$. Although adaptive algorithms modulate the participation rate based on market conditions, parametric cost models use the average realized participation rate over the duration of the order.

Below we present the DB models from older to newer and then we report on their performance using historical order samples.

1. Power Law Model

This is the earliest model, which generalizes the “square-root” class of models. It explains cost in terms of order size relative to ADV and daily volatility.

$$\bar{C} = b\sigma\left(\frac{Q}{V}\right)^{\alpha}. \tag{3}$$

The impact coefficient b and the impact exponent α are estimated by non-linear fitting on past order flow. Typical values for US markets are $b \approx 0.35$ and $\alpha \approx 0.40$.

2. Spread Cost Model

This is an extension to the power law model. It assumes that the total cost vs arrival has two components, spread cost (or slippage) and price impact (or mid-quote displacement). Spread cost is proportional to the bid-ask spread, and increases with participation rate. The price impact part is the same as in the power law model.

$$\bar{C} = \left(a + cr^{\beta}\right)s + b\sigma\left(\frac{Q}{V}\right)^{\alpha}. \tag{4}$$

Typical values for US markets coefficients are $a \approx -0.33, b \approx 0.31, c \approx 0.51$ and for exponents $\alpha \approx 0.30; \beta \approx 0.24$. This is the model that has been used in the Autobahn Equity post-trade transaction cost analysis in the US.

3. Trading Rate Model

This is the latest DB model. It assumes that the main explanatory variable for price impact is the participation rate (or normalized trading rate), since this is what is visible to the market. The model also has explicit dependence on the order duration. Shorter duration orders have smaller price impact than longer duration orders at the same participation rate. In addition to the price impact, the model contains a slippage term, proportional to the bid-ask spread.

$$\bar{C} = as + c\sigma\sqrt{T}\, r^{\beta}. \tag{5}$$

Dependence on order size is implicit. Order size over ADV, participation rate, and duration are linked via eq. (2). So the model could also be written in terms of order size and participation rate. Typical values for US markets are $a \approx 0.30$, $c \approx 1.05$ and $\beta \approx 0.65$. Note how this model reproduces the classical square root model in the limit $\beta = 1/2$ and after using eq. (2). We intend to begin using the trading rate model in the Autobahn Equity post-trade transaction cost analysis in the US.

Model Comparison

In this section we compare the above three models in terms of their ability to explain realized cost. We use a historical order sample of DB internal US order flow data from January through May 2018. About 500,000 eligible orders where analyzed. Distribution information on the relevant variables in our sample are shown in table 1.

	Arrival Cost (bps)	Order Size (% ADV)	Participation Rate (%)	Spread (bps)	Volatility (Ann %)	Duration (Fraction)
Min	-499.89	0.000	0.00	0.35	0.37	0.000
1st Qu.	-8.66	0.022	0.37	3.50	23.41	0.011
Median	1.79	0.081	1.03	6.19	31.24	0.080
Mean	5.39	0.295	3.38	11.10	35.79	0.296
3rd Qu.	18.96	0.263	3.22	12.31	42.96	0.618
Max	499.98	5.000	49.99	249.99	660.15	1.000

Table 1: Summary statistics for a representative sample of US order data collected over period Jan through May 2018

Orders are grouped by size over ADV and the model forecast and realized mean cost per group is plotted in figure 1.

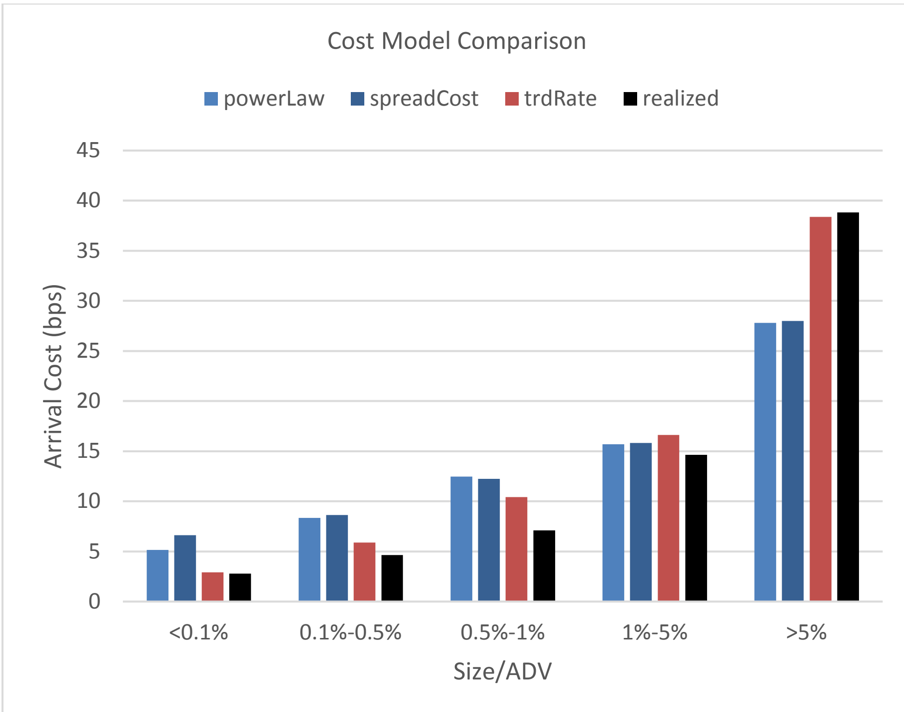

Figure 1 Model comparison across order sizes

It is clear that the trading rate model outperforms the other two in explaining realized cost across size bins. In particular, the model performs quite well for large size orders, i.e. greater than 5% ADV. The individual model forecasts are tabulated in table 2 below, along with the corresponding forecast errors.

It is also worth noting that the trading rate model contains no extra ad-hoc corrections or extrapolations for the high size/ADV range. The model is strictly defined by one functional form, as shown in equation (5). One could generalize the model by replacing the factor $\sigma\sqrt{T}$ with σT^{γ} and fit the extra exponent γ. This would introduce an extra scale to the problem, since the coefficient c will not be dimensionless any more. We prefer keeping the Brownian motion scaling $\sigma\sqrt{T}$. It seems to be adequate in capturing the fact that short duration orders are exposed to less price volatility and create less price impact for fixed participation rate.

Size/ADV	Order Count	powerLaw (bps)	powerLaw Error (bps)	spreadCost (bps)	spreadCost Error (bps)	trd Rate (bps)	trd Rate Error (bps)	realized (bps)
<0.1%	162,262	5.14	0.05	6.64	0.06	2.91	0.03	2.79
0.1%-0.5%	201,313	8.36	0.05	8.63	0.06	5.89	0.04	4.64
0.5%-1%	64,537	12.46	0.11	12.26	0.11	10.43	0.09	7.10
1%-5%	66,195	15.68	0.17	15.84	0.18	16.63	0.21	14.66
>5%	8,183	27.80	0.98	27.99	0.99	38.38	1.16	38.84

Table 2: Comparison of model forecasts and realized trading costs, grouped by order size/ADV

To further assess the choice of participation rate as a good explanatory variable, we break each order size group into four participation rate bins and plot the model and realized mean cost. The results are shown in figures 2 and 3.

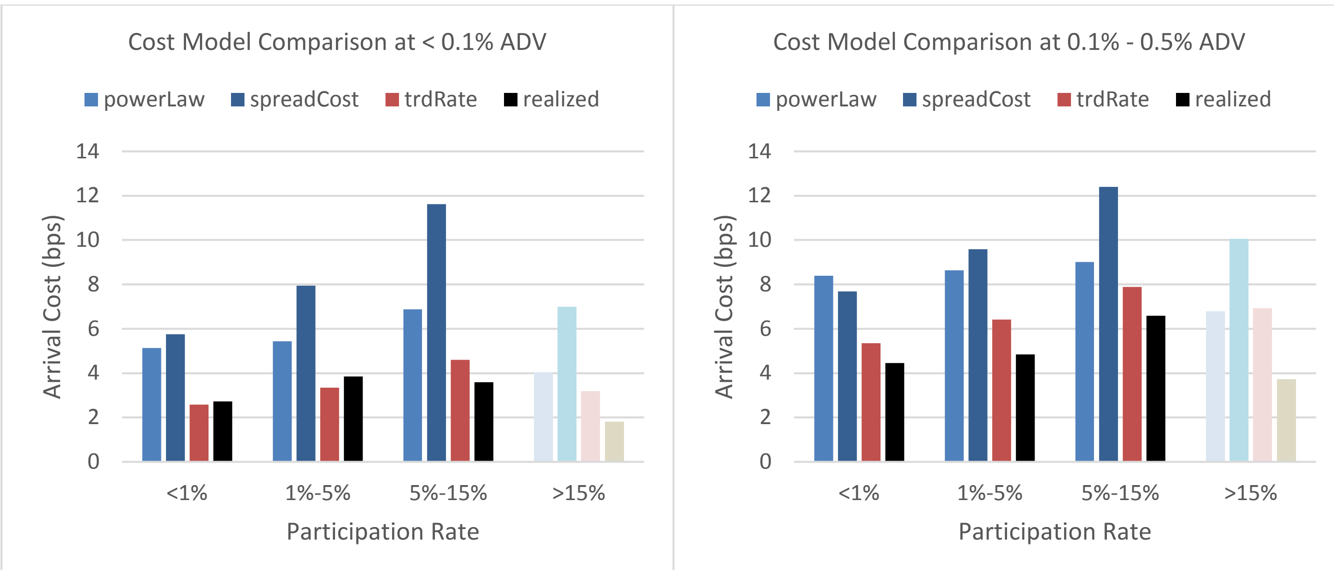

Figure 2 Average model and realized cost by participation rate for order sizes <0.1% ADV (left) and 0.1-0.5% of ADV (right)

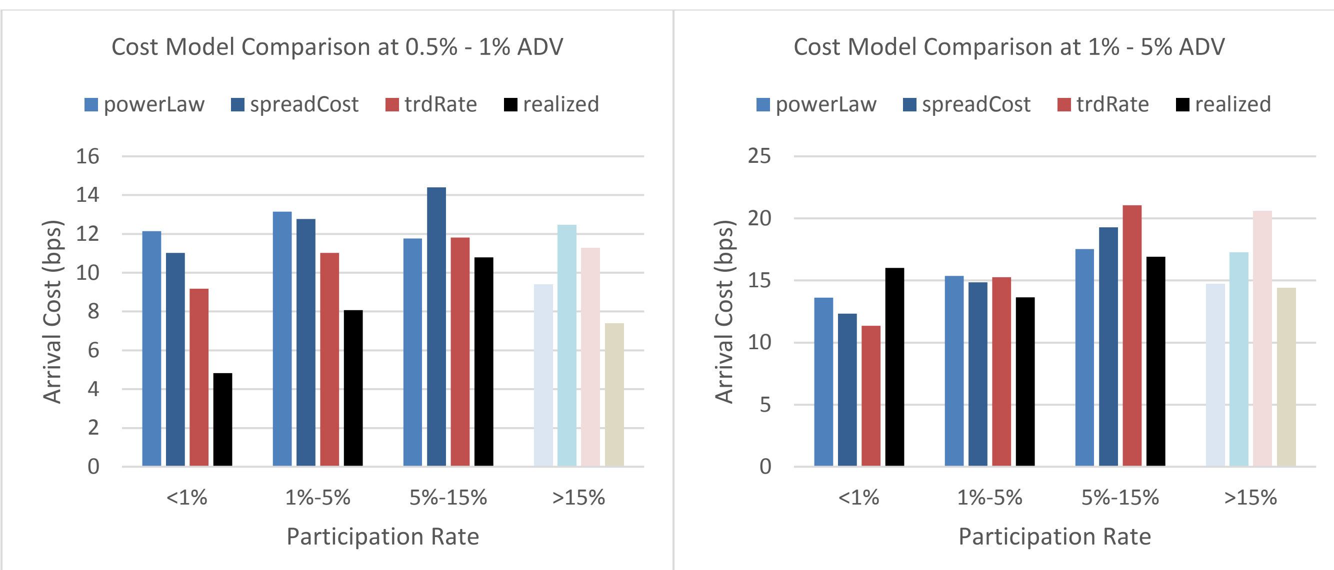

Figure 3 Average model and realized cost by participation rate for order sizes 0.5-1% ADV (left) and 1-.5% of ADV (right)

Statistical significance varies since the samples become thin at high participation rates. Nevertheless, the trading rate model stays closer to the realized cost than the other models, at least for order sizes up to 1% of ADV, where most of our sample concentrates. We expect to statistically strengthen the above results by analyzing larger samples in the future.

Conclusion

Parametric trading cost models explain the average cost of large order samples with a small number of relevant parameters connected via simple formulae. Their analytic structure makes them easy to work with in utility functions for optimization and trading strategy development.

There are some disadvantages in relying too heavily on parametric cost models to assess algorithmic performance, especially for smaller samples. Therefore, DB typically utilizes this metric as one reference point, in addition to several other factors, when evaluating post-trade performance and when undertaking experiments and algorithm parameter tuning.

Here we outlined the evolution of cost models developed and used at Deutsche Bank Global Equities. Calibration and refinement of the models is an ongoing effort. As such, we welcome and appreciate any feedback as this research effort evolves.

Contact

Michael G. Sotiropoulos, Ph.D.
U.S. Quantitative Research
michael.sotiropoulos@db.com | 212.250.5396

Garfield Brown, Ph.D.
U.S. Quantitative Research
garfield.brown@db.com | 212.250.1092

Alexandra Battle, CFA
U.S. Execution Consulting
alexandra.battle@db.com | 212.250.3988

Disclaimer