



 MarketAxess®

Composite+

ALGORITHMIC PRICING IN THE CORPORATE BOND MARKET

MARKETAXESS RESEARCH

David Krein | Global Head of Research

Julien Alexandre | Senior Research Analyst

Introduction

Composite+ (CP+) is MarketAxess' proprietary algorithmic pricing engine for corporate bonds. It is designed to support a variety of trading functions such as pre-trade price discovery, liquidity provision, transaction cost analysis, auto-execution, and crossing. It outputs an unbiased, two-sided market for more than 22,000 instruments globally. Updated every 15 to 60 seconds (depending on the liquidity of the instrument), the engine generates nearly 20 million levels per day covering 90-95% of trading activity in its markets.

In this paper, we will review CP+'s methodology, characteristics, and applications with a focus on TRACE-reportable US securities. This transparency provides more specific insight into the CP+ design and potential roles in traders' workflows.

CONSTRUCTION

CP+ works in two steps: (a) the prediction engine is "trained" each night using historical data; and (b) the engine generates its predictions in near real-time while incorporating all available intra-day information. All predictions are unique combinations of time, bond, and direction (meaning CP+ generates bids and offers independently).

The prediction engine rests on a supervised machine learning algorithm. The supervisory variable is the level (spread or price) of all qualifying reported client-to-dealer TRACE prints. Qualifications help narrow the subset of TRACE prints to those that best represent dealer risk pricing for institutional clients in a competitive market. For example, we limit consideration to institutionally-sized trades \$150,000 or larger, and exclude seemingly non-principal risk trades; e.g. paired transactions for which there is a corresponding client-to-dealer trade of the same size at roughly the same time in the opposite direction. These conditions are essential because they define the calibration of the algorithm.

The prediction engine employs three distinct sources of bond trading data: earlier TRACE prints, indicative bond price data streamed by dealers, and RFQ responses sent by liquidity providers via the MarketAxess trading platform. Additionally, it incorporates bond attribute information and other broad market data such as CDX levels.

We begin by creating a set of features, best defined as explanatory signals that aggregate the data from the past and the present to help predict the future. Each feature has a different predictive power that will depend on the vicinity to the prediction.

The most proximate features include TRACE prints, indicative levels, and RFQ responses from the same day on the same side for the bond of interest. This information is valuable in generating

the prediction, but the narrow constraints – a specific bond on a specific side on a specific day – may limit the coverage.

Additional subsets of features include progressively relaxed constraints. For example, features variously consider activity on the opposite side and in the inter-dealer market; activity within the last 90 days; and activity in other bonds with matching attributes. This last category allows features to incorporate activity in bonds with the same issuer, sector, rating and maturity.

Even when the most proximate signals are missing, the algorithm will learn to triangulate the less predictive features successively. This approach leverages all the available information to increase coverage without hurting quality.

Overall, the algorithm creates more than 200 features since it can make good use of even low information content factors to improve its prediction. The process also calls an outlier detection mechanism to identify and exclude off-market levels.

Each night, the features are fed into a tree-based machine learning algorithm called Gradient Boosting Method (GBM). GBM provides better accuracy than other algorithms like deep learning and neural networks since it is better at identifying the signal within noisy bond market data.

COVERAGE

COVERAGE	AVG # OF BONDS / DAY	% OF TRACE PRINTS*
U.S. Credit	12,248	93%

MARKET SEGMENT

Investment Grade	7,750	95%
High Yield	2,716	91%
Floater	784	84%
Short Corps	749	94%
Hybrids	248	97%

January 2018 data
Source: MarketAxess Research

*A TRACE print is covered by CP+ if we had a published CP+ level 15 minutes prior to the TRACE print. Only risk trades with a size greater than or equal to \$150,000 are considered.

BENCHMARKING

The CP+ algorithm – designed to predict, with minimal error, the level (spread or price) of qualified client-to-dealer TRACE prints that are \$150,000 or larger – can easily be tested out-of-sample with historical TRACE data.

Specifically, we benchmark TRACE prints above \$150,000 to the CP+ prediction 15 minutes prior. We use a 15 minute buffer because TRACE prints have to be reported within 15 minutes; further, when an RFQ is underway on MarketAxess, dealers can modify their inventory due to the new information they received from the existence of the inquiry. (This information shouldn't be taken into account when evaluating the accuracy.)

Further, we can benchmark those same TRACE prints to TRACE itself, using the most recent, or last, TRACE print up to 60 days prior with the same constraints.

We then look at 2 metrics:

Bias – Defined as the average of all distances (Δ) between the trade level and the CP+ prediction on the same side. A value of 0 means, that on average, we are correctly predicting the trade level. A positive bias means that, on average, the target trades wider than (outside away from the mid) CP+ and negative bias means that, on average, the target trades tighter than (inside towards the mid) CP+.

Error – Defined as the median of all absolute distances (Δ) above.

“Even when the most proximate signals are missing, the algorithm will learn to triangulate the less predictive features successively. This approach leverages all the available information to increase coverage without hurting quality.”

– DAVID KREIN, GLOBAL HEAD OF RESEARCH

BENCHMARK PERFORMANCE

	TRADE COUNT	CP+ BIAS (\$)	LAST BIAS (\$)	CP+ ERROR (\$)	LAST ERROR (\$)
--	-------------	---------------	----------------	----------------	-----------------

BY MARKET SEGMENT

Investment Grade	107,128	0.00	-0.04	0.08	0.15
High Yield	60,297	0.00	0.00	0.15	0.19
Floaters	5,550	0.00	0.00	0.04	0.05
Short Corps	14,611	0.00	0.00	0.02	0.03
Hybrids	3,881	0.00	0.00	0.20	0.25

BY LIQUIDITY SCORE

No score	636	0.04	-0.18	0.14	0.31
0-3	1,917	0.01	-0.01	0.19	0.28
4-5	31,052	0.00	-0.01	0.12	0.19
6-7	67,432	0.00	-0.01	0.10	0.14
8-10	90,430	0.00	-0.02	0.08	0.12

BY DEALER REPORTING SIDE

B	84,274	0.00	-0.02	0.09	0.13
S	107,193	0.00	-0.01	0.09	0.14

BY TRADE SIZE BUCKET

150k-1MM	114,895	0.00	-0.02	0.09	0.15
> 1MM	76,572	0.00	-0.01	0.09	0.12

January 2018 data

Source: MarketAxess Research

The universe is the set of TRACE prints that can be benchmarked to both CP+ and LAST (92% coverage).

APPLICATIONS

Access to CP+'s continuous two-way tick-level market data allows for new, or greatly improved, insights into the corporate bond market.

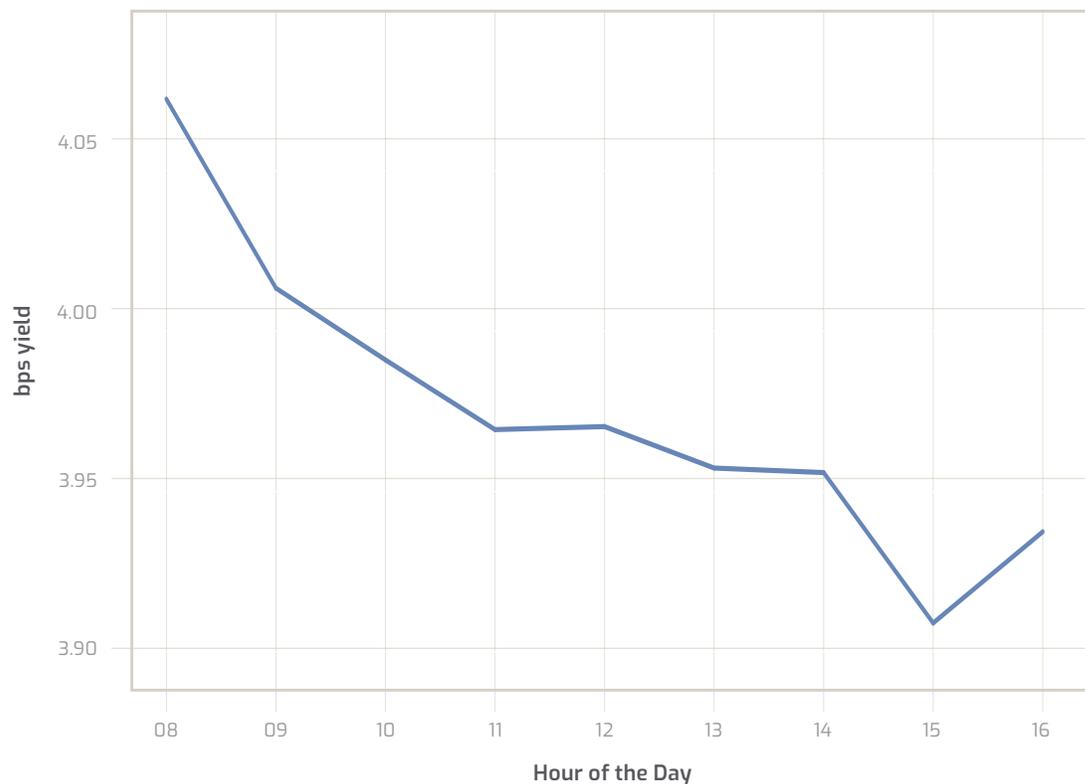
1. Intraday Bid-Ask Spread

The graph shows the hourly bid-ask spread for High Grade securities in January 2018 measured in basis point (yield) terms, i.e. bps. The universe of bonds is held constant on a daily basis. We see that the median decreases throughout the day as the trading information increases.

This is consistent with expectations. In early trading, there is little information available so the market, and by extension the CP+ model, reflect that risk with wider bid-ask spread. As the day goes by, more TRACE prints and RFQ responses come into play and we observe the bid-ask spread tighten accordingly.

The decline from high to low is roughly 0.13 bps.

MEDIAN BID/ASK SPREAD BY HOUR OF THE DAY



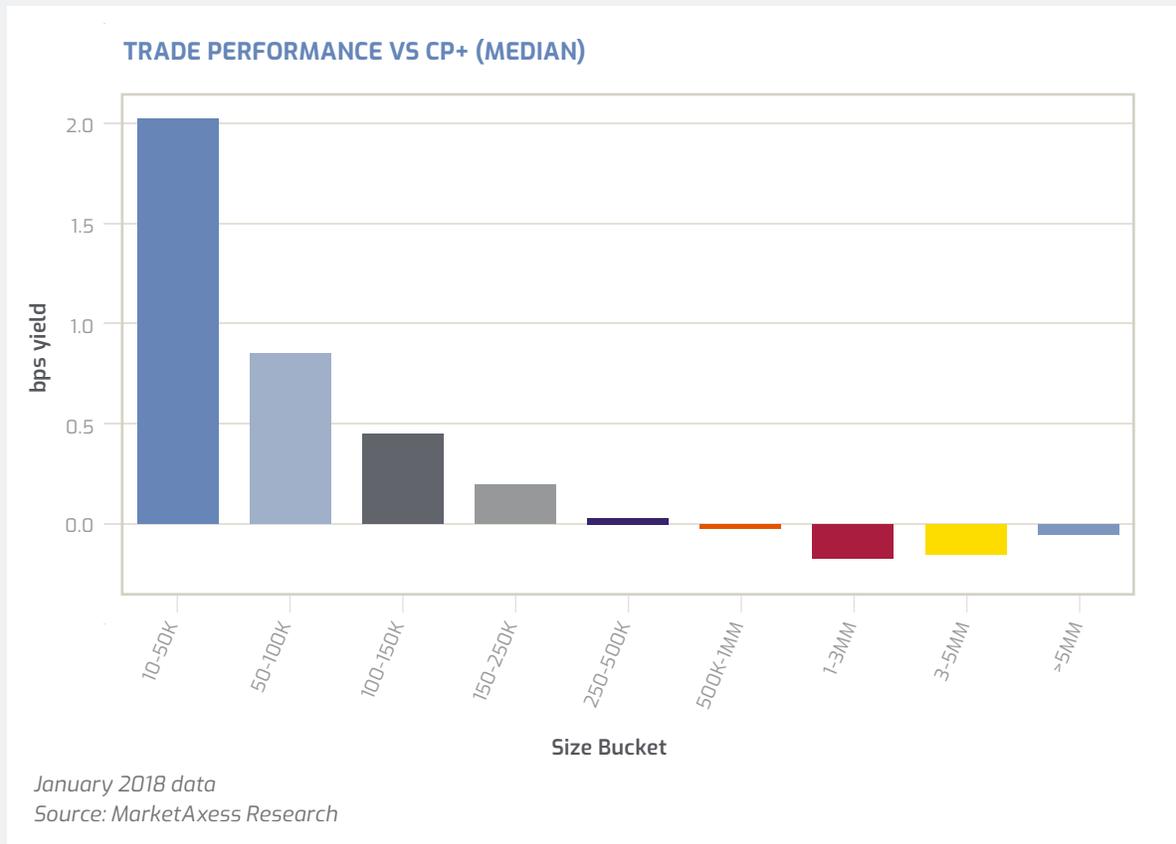
January 2018 data

Source: MarketAxess Research

2. Transaction Cost Analysis

The graph represents the US High Grade trade performance (in bps yield) by size bucket. It is defined as the distance between the TRACE trade spread (to benchmark) and the CP+ spread (to benchmark) on the corresponding side.

A negative value on the y-axis means that the trade was inside (better than) CP+. A positive value means that it was outside (worse than) CP+. Each size bucket's value in the chart is the median of all observations.



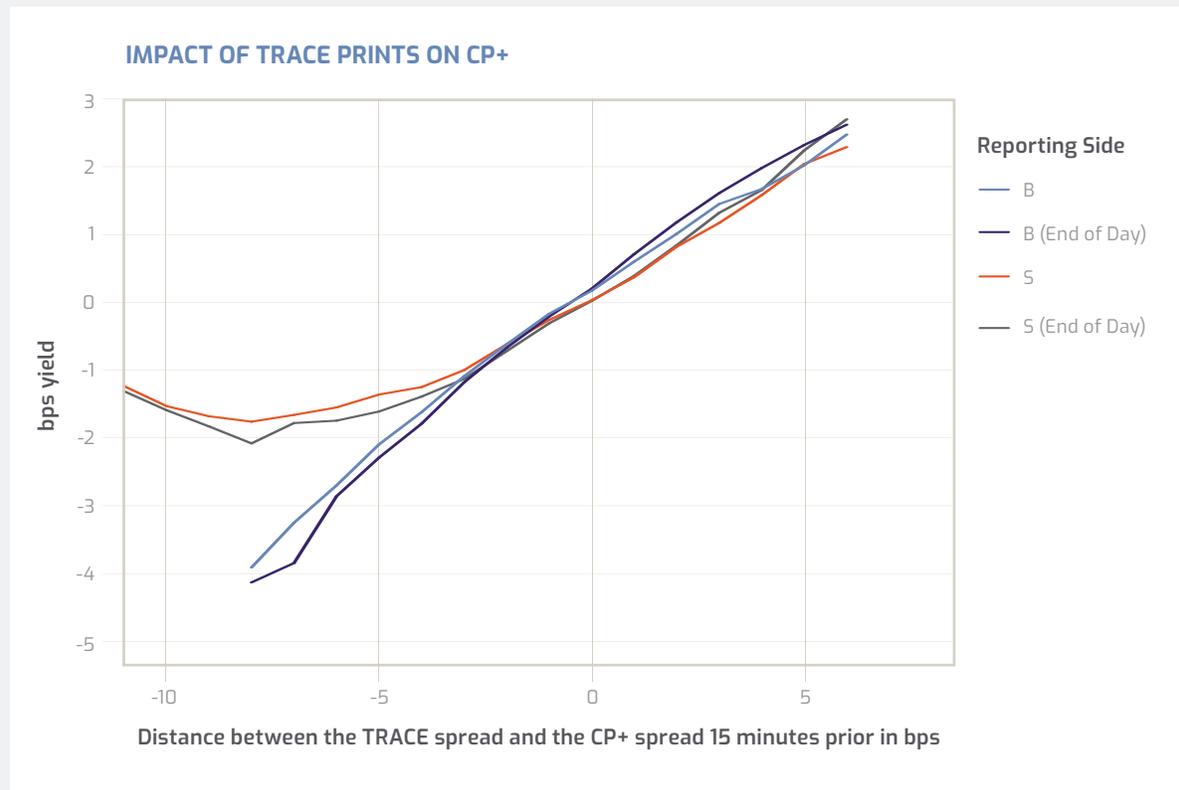
The graph proves that trades under 150k are priced wider than CP+ while trades above 150k are in-line with CP+. This confirms the assumption that institutional-sized trades start at 150k

and that calibrating the algorithm on these trades only, reduces the bias while including the maximum number of realized samples.

3. Price impact

Characterized as the movement of CP+ after a trade hits TRACE, the price impact graph below reveals the Bayesian nature of CP+.

On the x-axis, we have the distance between the TRACE spread (to benchmark) and the CP+ spread (to benchmark) 15 minutes prior. On the y-axis lays the price impact, defined as the median difference between the CP+ 15 minutes after the trade (or the CP+ at 4 pm for the End of Day version) and the CP+ 15 minutes prior.



A single trade can't be the only source that sets the price of a bond. Every other meaningful data point, even those less proximate to the prediction, will have an influence on CP+. Consequently, when a trade hits TRACE and deviates from CP+ by Δ , the impact on CP+ will be less than Δ . The algorithm still weights the other features when adjusting its price.

The impact is also side-dependent. 4.2% of Dealer Sells trades are at a high premium (10 bps or more below CP+ in yield terms), twice the share

of Dealer Buys trades (10 bps or more above CP+). The algorithm learns an interesting market effect: high premium to CP+ Dealer Sells trades seldom indicate a lasting price movement. CP+ moves by 1.5 basis point for a 10 basis points price difference, a very limited impact.

Furthermore, the immediate impact curve and the End of Day curve overlap, indicating that the algorithm neither over-reacts nor under-reacts.

Conclusion

- CP+ functions to transform data points that vary in quality and quantity, into an unbiased 2-way reference for more than 22,000 bonds.
 - By aggregating a broad and diverse set of data sources, including MarketAxess' proprietary data, it predicts subsequent trade levels with minimal error.
 - Its pertinent calibration on risk trades \$150,000 or larger empowers traders to adjust and decide.
-

For more information, please contact research@marketaxess.com

MarketAxess.com | ©2018 MarketAxess Holdings Inc. (the "Company"). MarketAxess, the MarketAxess logo, and "Now you're in the market" are trademarks of the Company. MarketAxess Corporation is a member of FINRA and SIPC. MarketAxess Europe Limited and MarketAxess Capital Limited are authorised and regulated by the UK Financial Conduct Authority. MarketAxess Singapore Pte. Limited is recognised by the Monetary Authority of Singapore. This information is not intended to represent an offer or solicitation of any financial instrument.

INFORMATIONAL PURPOSES ONLY. This material is for informational purposes only. It is not intended to form the basis of any investment decision, should not be considered a recommendation by MarketAxess or any other person and does not constitute an offer or solicitation to enter into a transaction with respect to the purchase or sale of any investment or security, nor is it a confirmation of terms. Sources of information are believed to be reliable, but MarketAxess makes no representation and gives no warranty that the information contained herein is accurate or complete. Any opinions or estimates given may change. MarketAxess undertakes no obligation to provide any **additional information or any update to or correction of the information contained herein.**