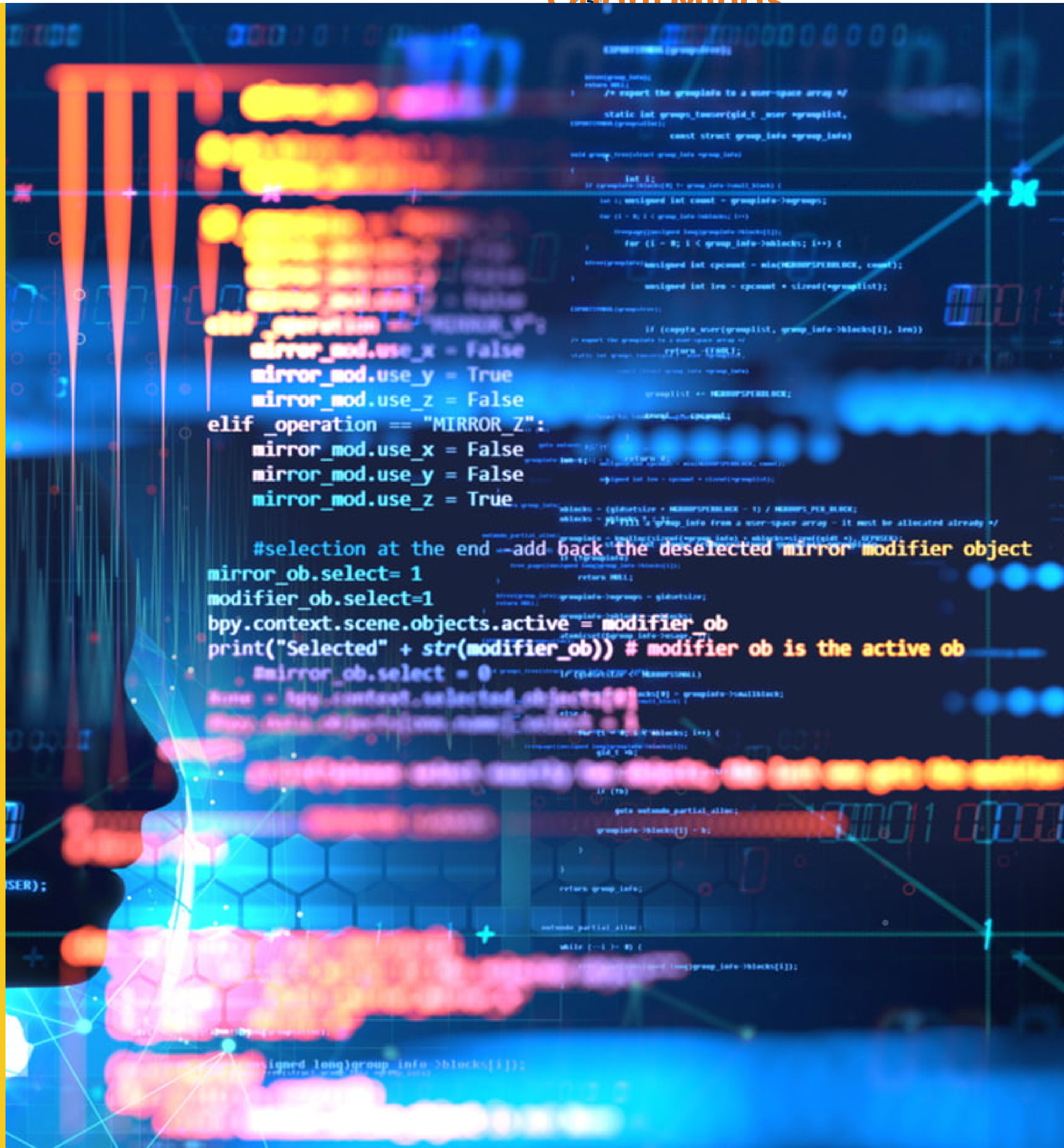
Innovations in quant trading strategies and modelling

Insights from QuantMinds International 2020



Contents & introduction

Contents

Introduction

Did you notice that the well-versed ways to make money no longer work? Old strategies fail because they're no longer relevant. People and markets have changed hugely, and quants need to adapt to this new equilibrium. Therefore, innovation, both technological and strategic, are in the spotlight in this QuantMinds eMagazine.

Exclusive in this edition, we catch up with Prado who is previewing of his new book. We learn more about neural networks with Alexandre Antonov, and start to understand and control their extrapolating behaviours. We explore with Jessica James what the changing landscape has to offer for quants, while Marcos Carreira shares his observations on the different market players. To top it off, Carol Alexander summarises some of her latest research into crypto assets and their microstructure.

We hope you will find this eMagazine just as insightful as we did. We're looking forward to learning more from these experts in May at the next QuantMinds International. See you there!

The QuantMinds Team

The future is bright... for those who survive

An interview with Marcos Lopez de Prado

Cornell University's Professor Marcos Lopez de Prado is a recognised authority in machine learning. In addition to his outstanding academic career, for the past 20 years he has managed multibillion-dollar funds at some of the largest and most successful asset managers. He founded True Positive Technologies in 2019, with the proceeds from his sale of several patents to AQR Capital Management, where he was a Partner and its first Head of Machine Learning. Cambridge University Press is about to release his new book, "Machine Learning for Asset Managers", and we asked him to give us a preview.

What is the main premise of your new book?

Whatever edge you aspire to gain in finance, it can only be justified in terms of someone else making a systematic mistake from which you benefit. Without a testable theory that explains your edge, the odds are that you do not have an edge at all. A historical simulation of an investment strategy's performance (backtest) is not a theory; it is a (likely unrealistic) simulation of a past that never happened. (You did not deploy that strategy years ago, that is why you are backtesting it!)

Only a theory can pin down the clear cause-effect mechanism that allows you to extract profits against the collective wisdom of the crowds; a testable theory that explains factual evidence as well as counterfactual cases (X implies Y, and the absence of Y implies the absence of X). Consequently, *asset managers should focus their efforts on researching theories, not backtesting trading rules*. Machine learning (ML) is a powerful tool for building financial theories, and the main goal of my new book is to introduce readers to essential techniques that they will need in that endeavour.

> “Asset managers should focus their efforts on researching theories, not backtesting trading rules
>
> Marcos Lopez de Prado

What's the worst mistake made by quantitative asset managers?

As I explained in my previous book, "Advances in Financial Machine Learning" (Wiley, 2018), backtesting is not a research tool. It is common for quantitative asset managers to confound research with backtesting. A backtest cannot prove a theory. A backtest only estimates how much a trading rule profits from an observed pattern, but it does not tell us whether the pattern is the result of signal or the result of noise. The strategy could be profiting from a statistical fluke. To answer that question, we need a theory that can be directly tested with greater depth than a mere historical simulation of a trading rule.

Academic journals are filled with papers where researchers backtest a strategy on decades (sometimes centuries!) of data, and present those results as evidence that a particular investment strategy works. Authors almost never control for selection bias, and in the absence of a theory we must assume that those findings are false, due to multiple testing.

How could this methodological mistake be avoided?

> Only a theory can establish the mechanism that causes the patterns that the strategy is presumably profiting from.
>
> Marcos Lopez de Prado

First, authors must state their theories in clear terms. A strategy is not a theory. A strategy is an algorithm for monetising the patterns that presumably arise from a theory. For example, consider a theory that partly explains volatility as the result of market makers widening their bid-ask spreads in response to imbalanced order flow. A strategy may buy straddles whenever the order flow becomes imbalanced. Even if the strategy is profitable according to a backtest, it does not prove that the patterns are due to signal. Only a theory can establish the mechanism that causes the patterns that the strategy is presumably profiting from. Testing the theory involves evaluating its ultimate and inescapable implications. Following the previous example, we could analyse FIX messages in search of evidence that market makers widen their bid-ask spreads in response to imbalaced order flow. We could also evaluate the profits of market makers who didn't widen their bid-ask spread under extreme order imbalance. Furthermore, we could survey market makers and ask them directly whether their response to imbalanced order flow is to withdraw from the market, and so on. In other words, testing the theory that justifies the strategy has little to do with backtesting. It has to do with the investigative task of defining the cause-effect mechanism.

Second, once the plausibility of a theory has been established, and only then, we should backtest the strategy proposed to monetise the theory. Remember, a backtest is merely a technique to assess the profitability of a trading rule. In the absence of that theory, a backtest is a data mining exercise that proves nothing. Surprisingly, much of the factor investing literature suffers from this lack of rigour. To this day, there is no strong theoretical justification for most factors, even though investors have poured hundreds of billions of dollars on them. ML can help build that economic rationale, as explained in my new book.

Quant-sceptics argue that statistical models are useless because they cannot predict black swans. What's your take?

Black-boxes cannot predict a black swan, because a model cannot predict an outcome that has never been observed before. Only a theory can do that. A theory must be general enough to explain particular cases, even if those cases are black swans. For instance, the existence of black holes was predicted by the theory of General Relativity more than five decades before the first one was observed. Black swans are extreme instances of everyday phenomena. In the earlier example, market microstructure theory explains how market makers react to order flow imbalance, leading to heightened volatility. The flash crash of May 6 2010 was a black swan, however its microstructure was predicted by the O'Hara-Easley PIN theory, going back to 1996. In conclusion, quantitative models are useful as long as they are supported by validated theories.

How does ML help uncover theories?

ML methods decouple the specification search from the variable search. What this means is that ML algorithms find what variables are involved in a phenomenon irrespective of the model's specification. Once we know which variables are important, we can formulate a theory that binds them.

This is an extremely powerful property that classical statistical methods (e.g., econometrics) lack. A *p*-value may be high for an important variable because the researchers assumed the wrong specification, leading to a false negative. Given how complex financial phenomena are, the chances that economists can guess a priori the right specification are slim. ML is the tool of choice in most scientific disciplines, and it is time economists modernise their empirical toolkit.

But aren't ML algorithms more prone to overfitting?

On the contrary, classical statistical methods are more likely to overfit, because they derive their estimation errors in-sample: The same observations used to train the model are also used to evaluate its accuracy. The reason for classical methods' reliance on in-sample error estimates is that these methods predate the advent of computers. In contrast, ML methods apply a variety of numerical approaches to prevent overfitting: cross-validation, regularisation, ensembles, etc.

Aren't financial datasets too short for ML applications?

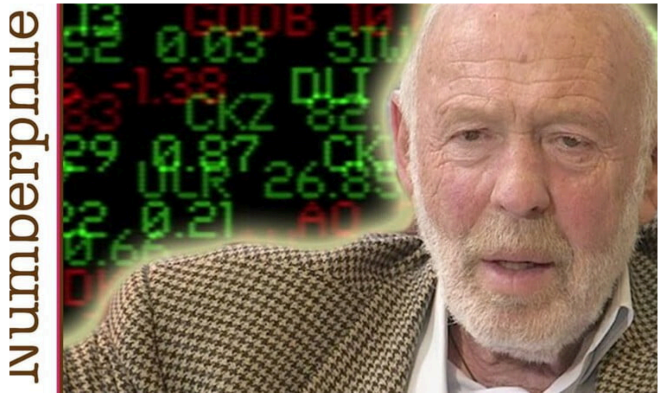
They may be too short for some deep neural networks, but there are plenty of ML algorithms that make a more effective use of the data than the classical statistical methods. For example, the random forest algorithm tends to perform better than logistic regression, even on small datasets, among several reasons because it is more robust to outliers and missing data.

What's your view on alternative data?

Twenty years ago, one could extract alpha using Excel, like most factor investment strategies attempt. You would rank descending stocks by P/E ratios, buy the bottom, and sell the top. Today, those strategies are mostly dead, as a result of crowding and backtest overfitting. If a strategy is so simple that anyone can implement it, why should anyone assume that there is any alpha left in that pattern?

Whatever alpha is left in the markets, it is more likely to come from the analysis of complex datasets, which require sophisticated ML techniques. I call this microscopic alpha. The good news is that microscopic alpha is much more abundant than macroscopic (unsophisticated) alpha ever was. One reason is, strategies that mine microscopic alpha are very specific (sometimes even security specific), which allows for an heterogenous set of uncorrelated strategies. Another reason is, firms chasing macroscopic alpha are a significant source of microscopic alpha, because the simplicity of their declared strategies makes their actions somewhat predictable. Accordingly, even if the individual Sharpe ratio of microscopic alpha were low, their combined Sharpe ratio can be very high.

RenTec is an example of a firm that has been successful at mining microscopic alpha with the help of ML, and continues to do so consistently, while traditional asset managers have failed to deliver macroscopic alpha. In short, alternative data is an important ingredient for success, in combination with ML and supercomputing.



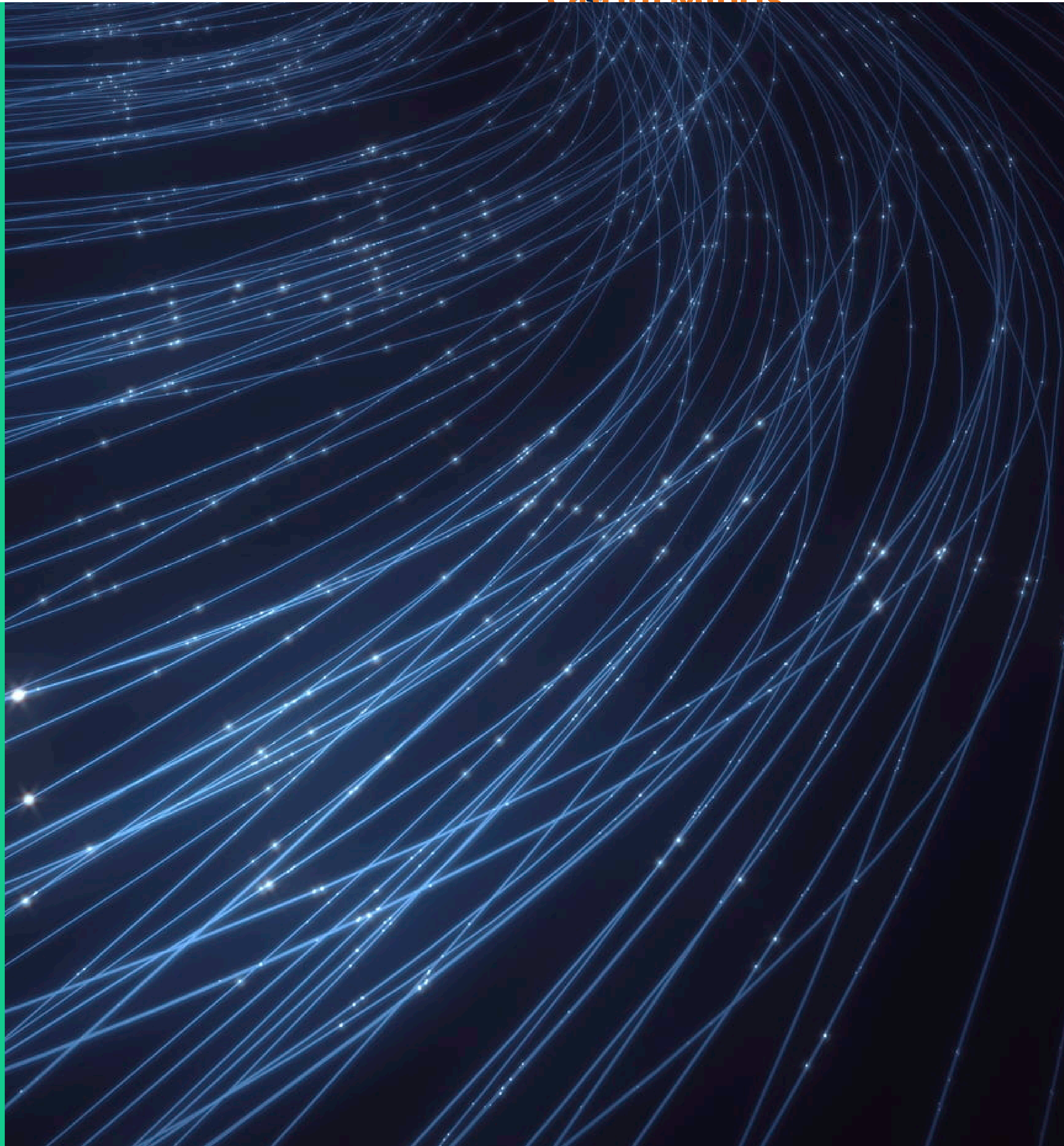
You founded True Positive Technologies (TPT) last year. How does TPT help investors?

TPT helps bring asset managers into the Age of AI. We develop customised investment algorithms for institutional investors. My partners and I founded TPT by popular demand. In less than one year, we have been engaged by firms with a combined AUM that exceeds $1 trillion. This reception has surpassed our wildest expectations, so we are very pleased with the industry's desire to modernise.

I'm more concerned about the state of affairs in academia. Economics students should be exposed to modern statistical methods, following the trail of students from other disciplines. The study of basic econometrics should be complemented with advanced courses in ML. The complexity of alternative datasets is beyond the grasp of econometrics, and I fear that students are only being trained to model (mostly irrelevant) structured data.

Neural networks with asymptotics control

How can our knowledge of asymptotics translate into control over the extrapolating behaviour of NNs?



Applications of artificial neural networks in quant finance have met various challenges, one of which is our current inability to control the extrapolation behaviour of NNs beyond the range of training points. In the working paper "Neural Networks with Asymptotics Control", Alexandre Antonov, Michael Konikov, and Vladimir Piterbarg demonstrate how the knowledge of asymptotics can be translated into control over the extrapolating behaviour of NNs, and in this article, Alexandre Antonov, Chief Analyst, Danske Bank, explores the key concepts supporting this paper.

Significant advances in machine learning (ML), deep learning (DL), and artificial neural networks (ANN or NN) in image and speech recognition fuelled a rush of investigations as to how these techniques could be applied in finance in general, and in derivatives pricing in particular. Typical examples of this genre include [AK], [MG], [HMT], [FG].

The main idea of these papers is to use NNs to speed up slow function calculations. A typical procedure involves training the NN offline on a sample of learning points calculated from the true model (or, in pre-NN language, fitting a functional form defined by an NN to a sample of function values over a collection of function arguments, often multi-dimensional), and then using the NN as an approximation to the true model during on-line pricing and risk management calculations.

It has generally been observed that NNs, once trained, do a good job interpolating between the points they were trained on (fitted to). However, extrapolation behaviour beyond the range of training points is not controllable in a typical NN, due to their complex non-parametric nature. In this paper, we provide a detailed description on this absence of extrapolation/ asymptotics control. Starting with an extensive intuition on the Kolmogorov-Arnold theorem underlying a standard feed-forward multi-layer NN, we demonstrate how an information about asymptotics is lost.

The absence of extrapolation control is a significant limitation of the NN approximation approach to financial applications. The obvious one is stress testing. Financial models often need to be evaluated with the values of input variables that are significantly different from the current market conditions. Changes of regime are common in financial markets. Moreover, input values in stress scenarios, required for sound risk management, would routinely fall outside the range of the training set, with unpredictable extrapolation. Needless to say, there are many other reasons why it is important to control extrapolation of NNs in financial applications.

> "
> Large ranges of input values need a large number of learning points to cover, slowing down learning
>
> Alexandre Antonov

One of the possible solutions to this problem is, of course, sampling the input variable space widely enough so that any possible future value of input variables falls within the sample range (interpolation) and never outside (extrapolation). It is not hard to see that this is not a fully satisfactory solution as one does not know a-priori what future values will be required. Additionally, large ranges of input values need a large number of learning points to cover, slowing down learning. More importantly, using large ranges for input variables would likely make the fit for moderate, i.e. non-extreme, values of inputs worse, as the NN would try to balance the quality of fit between all the training points.

Fortunately, for many financial applications, in addition to the ability to calculate function values for moderate values of inputs, we also often know asymptotics of these functions for large values of parameters. This is true for e.g. SABR fitting ([MG], [HMT]) and values of many types of products in derivatives pricing ([FG]). The aim of this paper is to demonstrate how the knowledge of asymptotics can be effectively translated into control over the extrapolating behaviour of NNs.

Namely, to approximate a multi-dimensional function while preserving its asymptotics we come up with two steps. As the first one, we find a control variate function that has the same asymptotics as the initial function. On step two, we approximate the residual function with a special NN that has vanishing asymptotics in all, or some, directions.

The apparent simplicity of the plan hides a number of complications that we overcome in this paper. Specifically, we make two critical contributions – our main technical results -- that make this programme work. For step one, we show how to construct a universal control variate, a multi-dimensional spline that has the same asymptotics as the initial function. For step two, we design a custom NN layer that guarantees zero asymptotics in all directions, with a fine control over the regions where the NN interpolation is used and where the asymptotics kick in.

In passing, we note that multi-dimensional interpolation is not the only application of NNs in quantitative finance; papers [KS], [BGTW], [GR], [HL] explore other applications that are beyond the scope of this paper. Still, they may benefit from some of the ideas presented here.

More details as well as intuitions and multiple illustrations can be found in our SSRN paper:A. Antonov, M. Konikov and V. Piterbarg, "Neural Networks with Asymptotics Control", 2020, SSRN working paper

References

[BGTW] H. Buehler, L. Gonon, J. Teichmann, B. Wood. "Deep Hedging", 2018, arxiv. Control Signal Systems (1989) 2: 303.

[FG] R. Ferguson and A.D. Green, "Deeply Learning Derivatives", 2018, SSRN working paper

[HL] P. Henry-Labordere. "CVA and IM: Welcome to the Machine", March 2019, Risk Magazine

[HMT] B. Horvath, A. Muguruza and M. Tomas, ``Deep Learning Volatility'' 2019, SSRN working paper

[AK] A. Kondratyev, "Curve dynamics with artificial neural networks", 2018, Risk

[KS] A. Kondratyev, C. Schwarz. "The Market Generator", 2019, SSRN working paper,

[MG] W. McGhee, "An Artificial Neural Network Representation of the SABR Stochastic Volatility Model" (2018), SSRN working paper

[GR] G. Ritter. "Machine learning for trading". October 2017, Risk Magazine, pp.~84-89

Can FX hedges of bonds deliver a free lunch?

Under just the right circumstances, the elusive free lunch may temporarily appear on the table

We all know that in the markets, there is no such thing as a free lunch. Risk-free returns don't exist – and if they did, even briefly, they would be traded on until they disappeared. Except sometimes, somehow, under just the right circumstances, the elusive free lunch may temporarily appear on the table... In this article, Jessica James, Managing Director, Senior Quant Researcher, Commerzbank, explores how.

I would like to talk about government bond yield differentials. At the time of writing, US 2y government bonds deliver about 1.5%. Similar EU bonds are giving -0.6%. Investors in high grade debt based in Europe are enviously glancing over the pond.

Why don't they sell some EU bonds and buy the US bonds? Ah, says your Finance 101 lecturer, because of the FX risk. If a EUR based investor buys a US bond, the FX rates can swing by several percent in a few days or weeks – wiping out the carefully calculated interest rate differential. The FX risk overwhelms the trade.

Well then, why does the investor not hedge the FX risk of the trade? Once more, the Finance 101 lecture notes have an answer – because the cost of hedging will exactly offset the interest rate differential. So our investor sadly regards his underperforming portfolio and wonders what to do. But! His notes are from pre 2008 days – and these days, things are a little different.

> The FX hedge in the post crisis world is affected by the cross currency basis, which means that sometimes, the FX hedged deal is a lot more interesting.
>
> Jessica James

Maturity-matched hedges

As a EUR based investor, the higher yields currently available for US bonds across the tenor range are attractive. How could the EUR based investor take advantage of this?

In a financial market with no capital charges or regulatory and XVA issues, then it would not be possible to preserve any kind of yield pick-up by investing in a foreign bond and then hedging the FX risk. The 'theoretical' forward FX rate would, in the pre-2008 crisis days, have almost exactly cancelled out the yield pick-up, though there might have been some credit spread available. But even this was small.

Now, however, there exists in general a substantial cross-currency basis, meaning that the arbitrage relationship between FX spot and forward, and interest rates, has broken down to an extent, due partly to demand for different currencies and partly to regulatory activity which restricts arbitrage trades. The actual cross currency basis is usually expressed as a difference to the non-USD interest rate of the currency pair, though it could equally well be expressed as a spread to the USD rate or to the FX spot or theoretical forward rate. Additionally, credit spread differentials have become far more significant since the crisis, and may provide additional pick-up in different industries and tenors.

We would execute the hedge in the swap market, which has slightly different rates from the government bonds (though they are correlated), and we would include the basis swap. The (US) basis swap is the extra cost of borrowing US dollars via a currency swap compared to what it should be purely according to interest rate differentials.

What, then, will the cost of hedging be?

The mechanics of the package are thus: the EUR based investor exchanges EUR for USD, buys the USD bond, and puts on a maturity-matched FX hedge. Simplistically, assuming a 1-year period, and without worrying about coupon payments, we can write this as below.

FX is the FX rate at the start of the deal

FR_1 is the 1 year forward FX rate for the bond maturity

$$P = \text{EUR principal amount at start of deal}$$

$$P \times FX = \text{USD principal amount at start of deal}$$

$$P \times \frac{FX}{FR_1} = \text{EUR principal amount at end of deal}$$

As said, we are not worrying about hedging the interest rate risk. Thus

$$\text{EUR hedge cost } = P - P \times \frac{FX}{FR_1} \tag{1}$$

But from the arbitrage based construction of the forward rate FR, we know that

$$FX = FR_1 \frac{1 + R_1}{1 + R_2}$$

Where R_1 is the EUR interest rate, and R_2 is the USD interest rate for the bond tenor.

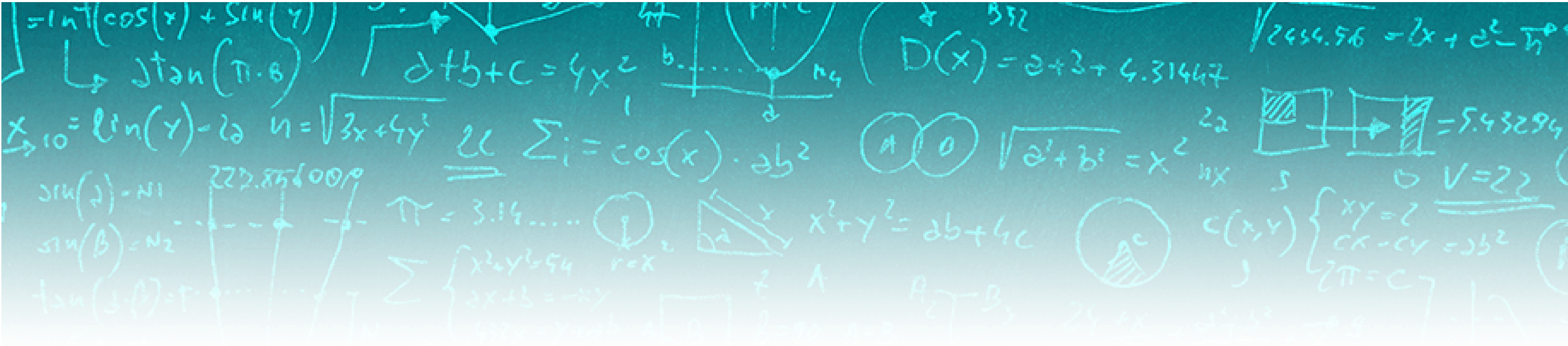
Thus EUR hedge cost is given by

$$\text{EUR hedge cost} = P - \left[\frac{P}{FR_1}\right]\left[FR_1 \frac{1 + R_1}{1 + R_2}\right]$$

$$\text{EUR hedge cost} = P\left[1 - \frac{1 + R_1}{1 + R_2}\right] = P\left[\frac{R_2 - R_1}{1 + R_2}\right] \approx P[R_2 - R_1]$$

R_2 is the USD swap rate, but we actually now need to adjust it by the basis swap amount. Thus

$$\text{EUR hedge cost} \approx P[R_2 + R_{basis} - R_1] \tag{2}$$



Because (1) and (2) are approximately equal to each other, it can be calculated either way, depending on the data which is available. One would usually use equation (2) as the basis swap is explicitly incorporated, but if one wished to use the spot and forward FX rates, then it is possible to express the quantity [$R_2 + R_{basis} - R_1$] in terms of these FX rates as in equation (1), because the forward FX rate in the market does incorporate the basis.

It's not difficult to extend this to the multi-year case. We know that for n years,

$$FX = FR_n \frac{(1+R_1)^n}{(1+R_2)^n}$$

So the total EUR hedge cost over the whole deal is given by

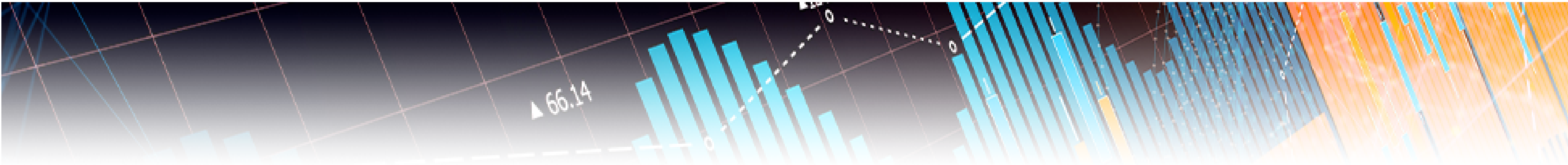
$$\text{EUR hedge cost} = P\left[1 - \frac{(1+R_1)^n}{(1+R_2)^n}\right] = P\left[\frac{n(R_2 - R_1)}{1 + nR_2}\right] \approx P \times n[R_2 - R_1]$$

$$\text{EUR annual hedge cost} \approx P[R_2 - R_1] \approx P\left(1 - \left[\frac{FX}{FR_n}\right]\right)/n$$

...if we use a binomial expansion and take only the first order. Once more of course we actually need to include the basis swap so the actual annual cost will be $P[R_2 + R_{basis} - R_1]$.

$$\text{EUR annual hedge cost} \approx P[R_2 + R_{basis} - R_1] \approx P\left(1 - \left[\frac{FX}{FR_n}\right]\right)/n \qquad (3)$$

So we may now calculate a hedged yield pick-up which is actually accessible to the EUR based investor. It is FX hedged and relatively risk free. Note that equation (3), where the hedge cost is expressed in terms of the FX spot and forward rates, is a novel way of calculating hedge cost. It is useful in that it is an alternative way of arriving at the answer from other data series than those usually used.



Practical calculation of maturity-matched yield pick-up

From now on, we can omit the principal amount P in expressions as it will always cancel on both sides of the equation and there is no loss of generality in expressing all quantities as percentages.

Once we have the hedge cost, we may derive the annualised yield pick-up very simply, as below

$$\text{Yield pick} - \text{up} = [B_2 - B_1] - \text{Hedge Cost}$$

As previously discussed, in a perfectly efficient market, this would be zero, but in the 'real world' it is often substantial.

In practice, we may derive this yield pick-up three ways

(1) Using bond yields, swap rates, and cross-currency basis swaps

$$\text{Yield pick} - \text{up} = [B_2 - B_1] - [R_2 + R_{basis} - R_1]$$

(2) Using asset swap spreads, 3s6s basis swaps, and cross-currency basis swaps

$$\text{Yield pick} - \text{up} = [A_1 - A_2] - C_1 + R_{basis}$$

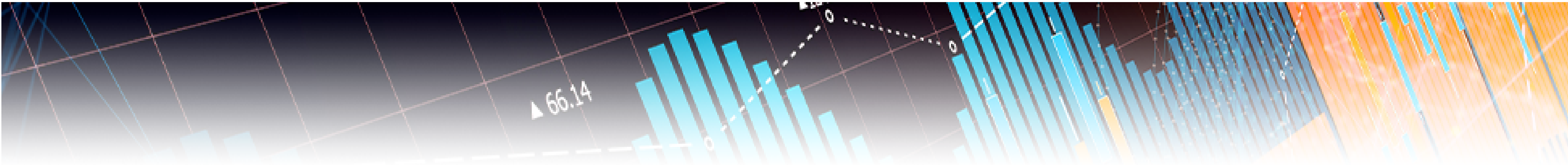
(3) Using bond yields and spot and forward FX rates

$$\text{Yield pick} - \text{up} = [B_2 - B_1] - \left(1 - \left[\frac{FX}{FR_n}\right]\right) / n \tag{4}$$

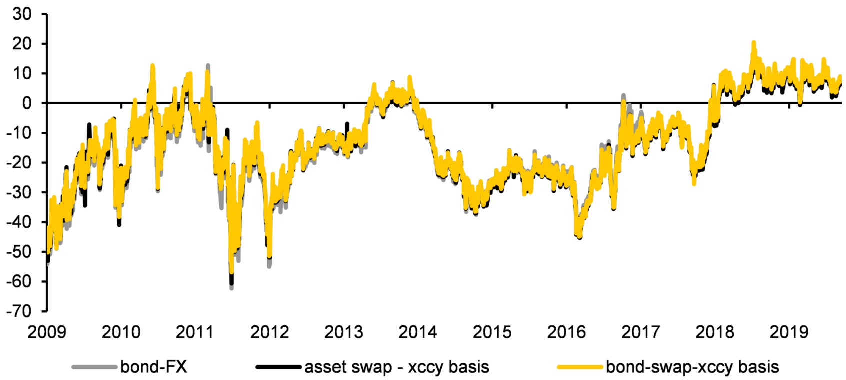
Where, for a EUR investor buying a USD government bond,

B_1 = EUR bond yield
B_2 = USD bond yield
R_1 = EUR swap rate
R_2 = USD swap rate
R_{basis} = XCCY basis swap
A_1 = EUR asset swap
A_2 = USD asset swap
C_1 = 3s6s basis
FX = spot foreign exchange rate
FR_n = forward foreign exchange rate for tenor n
n = tenor in years

Note that the 3s6s basis (ie., the difference between swap rates referenced to 3m and 6m Libor) may sometimes be needed if the other interest rates differ in their coupon frequency, and that the asset swap spread is A_1 - A_2 rather than A_2 - A_1 because it is always quoted as a positive spread over the government bond. Additionally, one may need to be careful of the sign of the basis swap which is usually quoted as a spread to the non-USD interest rate, so if the basis-adjusted EUR interest rate is lower than the actual rate, then the basis swap will be a negative number.



Having done all of this, we see in the graph below that the three ways of calculating the yield pick-up match beautifully!



Maturity-matched FX hedged yield pick-up for 2y USD bonds vs EUR vs German bonds, in bp | Source: Commerzbank Research, Bloomberg

It's obvious that the 'pick-up' is not always positive – the point one can make is that a negative pick-up from the perspective of one currency of a pair is a positive one from the other currency's perspective. But the fact that at times, certain investors have been able to lock in a pick-up of 50 bp or more does make one think that just occasionally, lunch comes without a bill.

Now, this is of course not the whole story. We've assumed that the investor can short EUR government bonds in the repo market without cost, which is not true. And these pickups are only strictly available if the investor holds the bonds until maturity – two years, in the above example. This does not suit the majority of portfolio managers. However, it is good to think that in today's world, many of the old certainties which are no longer true give way to far more interesting situations.

References

[1] For an in-depth discussion of what the basis is, see Commerzbank's Rates Radar "More interest in cross-currency basis swaps", March 2017

The right kind of volatility

How do market makers make money?

Years ago, Rebonato was discussing the 2007 Quant Apocalypse (I think), and he paraphrased the infamous "the wrong type of snow" from a British Rail interview to say that what happened was "the wrong kind of volatility". All the time we hear that proprietary trading firms make money with volatility, and that periods with low volatility are bad for them. So, Marcos Costa Santos Carreira, PhD Candidate, École Polytechnique, investigates: is there a right kind of volatility for market makers? And how do market makers make money?

For that we will have to look at how markets work and try to model some of these features, starting with a contract where prices trade at certain prices only; the difference between these values is known as the tick value (e.g. 0.01 for US equities); so prices are multiples of the tick value (represented in our papers as α). Those not familiar with this subject might have heard about the decimalisation in US equities at the end of the 90s, where 1/8s and 1/16s were dropped in favor of the 0.01 price increment.

This discretisation means that what we observe (repeated trades at the same price, discrete price changes and the times between these changes) is representing some kind of hidden process, where agents make decisions based on variables like the order book imbalance (i.e. whether the size of the order book on the first (top) level(s) is bigger on the bid (ask) side than on the ask (bid) side.

So let's assume that there are three types of agents: (i) market makers, who try to post liquidity on the top of each queue to capture the spread as (ii) noise traders cross the spread without depleting the existing liquidity at the price level traded; the traders who deplete the existing liquidity on one side of the order book are (iii) informed traders. We are adapting the definition from the paper [From Glosten-Milgrom to the whole limit order book and applications to financial regulation]().

"As the tick value gets lower, more trades become informed; as the tick value gets higher, more trades become noise. But what else contributes to it?

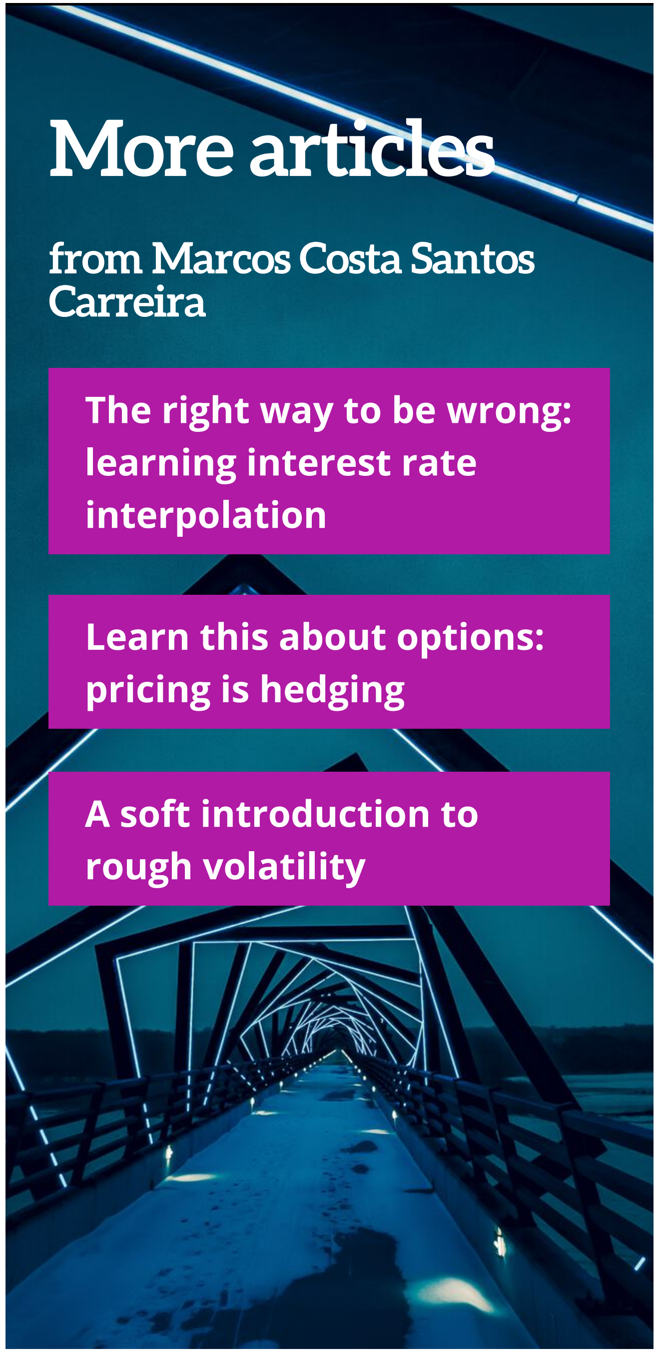
Marcos Costa Santos Carreira

As the tick value gets lower, more trades become informed; as the tick value gets higher, more trades become noise. But what else contributes to it?

Well, volatility should play a role; after all, the more volatile an asset is, the more the price will change, leading to faster price changes and a wider range of prices over a trading day.

We have already looked at the relationship between volatility and tick sizes in the paper A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones and 3 years ago in the presentation Microstructure Of A Central Limit Order Book In FX Futures; we know that the number of price changes in a trading day is proportional to the square of the volatility and to the square of the inverse of the relative tick value (α divided by the asset price). We also know that the parameter η, defined as ½ multiplied by the ratio between the number of continuations and the number of alternations (where continuations refer to consecutive price changes with the same sign and alternations refer to consecutive price changes with opposite signs) plays a role in the number of trades.

We've been conducting further research with the help of the CME, and looking at the decreasing volatility of FX future contracts we started to notice that the number of trades and the trade durations (times between price changes) seemed different than expected; the answer is that, at the end of the day, prices will not be the same as of the opening; so a better way to model the evolution of the hidden price process is to look at the effect of both the volatility and the trend of the process; remember that usually in financial mathematics the time t multiplies the trend μ but also the square of the volatility σ; so a big reduction in σ makes the trend more important; so when counting the number of price changes one will find more continuations than expected.

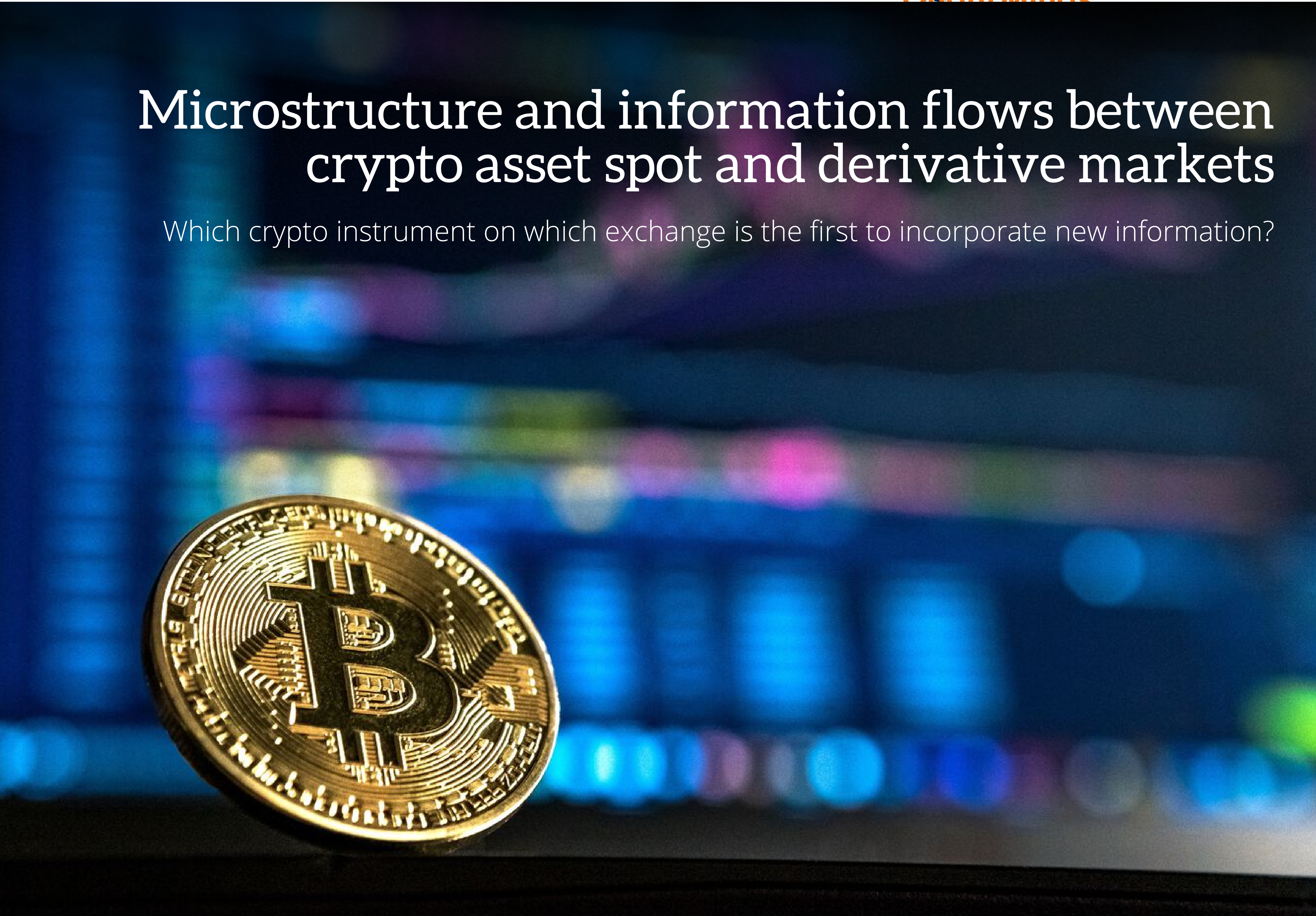


Our new formulas can account for this difference, and more important we can now understand better the role that volatility and trend play. For a market maker, the volatility σ is good, as it leads to alternations and the opportunity to earn a fraction of the spread on average; but the trend μ is bad for the simple strategy of posting liquidity in the top of the book; so, following some of the ideas of From Glosten-Milgrom..., we can think about the value of being on the top of the queue as something close to (1-2η)/(1+2η) on average; but because η can become quite large if the price moves relentlessly in one direction, we can see that there is no value in being a market maker in this situation, and market makers might become market takers. The parallels between the relative roles of the trend μ and the volatility σ and the ratio between informed trades and noise trades are stricking (in fact, we will argue that the ratio r on From Glosten-Milgrom... should be equal to 2η), and that trying to infer the parameter η from the price changes and rice durations is equivalent to infer the recent values of μ and σ, which is equivalent to trying to infer the proportion of informed trading. More details will be discussed in the presentation at QuantMinds International.

We hope that this model will be of interest to regulators and exchanges. For regulators, using this approach might lead to a faster diagnostic that continuous trading might be failing to provide adequate liquidity and price discovery (as the trend itself becomes the information) and an auction might be reasonable when sudden moves happen; this would be better than to wait for a 5% or 10% price move. For exchanges, it would help them not only to better choose the appropriate tick values but to monitor which factors (volatile volatilities, relative presence of informed traders like institutional investors, etc.) are in play. And it helps one to understand that, because the price changes are a consequence of both μ and σ, that not every volatility is the right kind of volatility for market makers.

Microstructure and information flows between crypto asset spot and derivative markets

Which crypto instrument on which exchange is the first to incorporate new information?



Professor of Finance at the University of Sussex Business School and leader of the Quantitative Finance and Fintech research group, Carol Alexander is presenting some path-breaking results at QuantMinds International 2020 – starting with the complex task of analysing price discovery among the hundreds of crypto spot and derivatives platforms which, for the past year, have had an average aggregate monthly trading volume of well over $500 billion[1].

Which crypto instrument on which exchange is the first to incorporate new information? Where are the most informed traders located? How long do traders on other platforms have to profit from the leaders reaction to news?

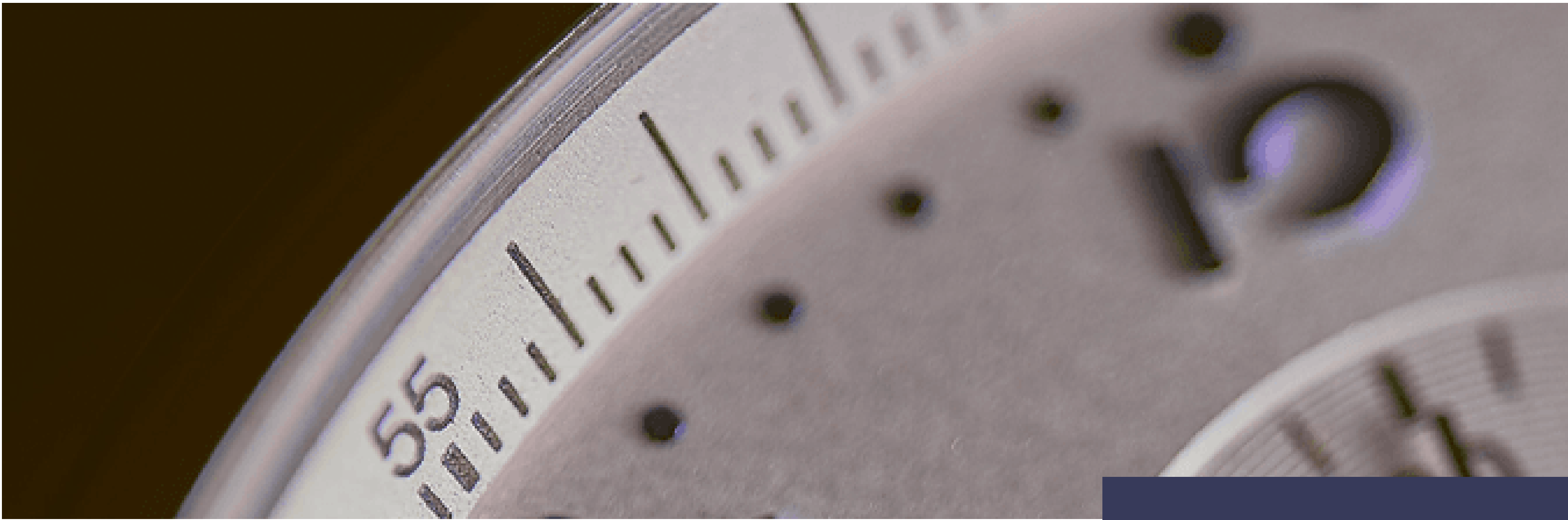
Since the 1980s, a massive amount of academic research on these questions aims to provide informational advantages to high-frequency traders. They are relatively easy to answer for traditional assets like financial futures/ETFs, gold, fiat currencies, energy and commodities. Compared with crypto, the trading is slower, the volatility lower, and there are only a few instrument-types in each asset class, typically traded on just one venue.

However, bitcoin and other cryptographic coins and tokens form a new class of financial asset with some very novel and interesting properties, including the market microstructure. They may be traded 24/7 and 365 days per year, either 'off-chain' on hundreds of different centralised exchanges or transferred directly 'on chain' using smart contracts on the Ethereum chain or other public blockchains via peer-to-peer (P2P) trading venues called decentralised exchanges.

Few of these venues are regulated, yet a wealth of limit order and transaction data are freely available, providing a level of transparency that is quite unique. As a result, crypto assets and their derivatives have become the ideal testing ground for studying market microstructure and price discovery, all the way from leader identification to optimal trade execution based on machine-learning algorithms.

About the data. For decentralised exchanges, block explorers give details of each transaction in every block, pseudo-anonymised with hexadecimal numbers representing wallet addresses in the P2P network. For centralised exchanges, although nothing is recorded on the chain[2], there is even more data. A wealth of free VWAP coin price and cap-weighted crypto market indices are freely available, although some are better quality than others[3].

Also, most platforms have an API which allows order book data to be downloaded directly at very high frequency. Data and software provider **CoinAPI** links with hundreds of crypto spot and derivatives exchanges, offering historical and streaming order book and trades in tick-by-tick or highest granularity data from all major centralised (and decentralised) exchanges.



Information share metrics for price discovery are derived from N-dimensional vector error correction models (VECMs) based on very high-frequency time series. Each day, my PhD student **Daniel Heck** has calibrated such a model to N different minute-by-minute returns and then calculated the metrics for each market on that day. The sum of the N individual information metrics is 1 and the higher the metric for a particular market, the more influential its price discovery role on that day. Then we roll forward 24 hours and repeat, getting a picture of the influence that different markets have on 'discovering' the common efficient bitcoin price, and how this leadership evolves over a period of time.

Here I'll focus on the metric called the 'generalised information share' (GIS). I'll be discussing the latest results for a selection of VECMs during the first part of [my talk at QuantMinds]().

But in this article I'll present just two VECMs, with N = 2 and 8 respectively: The first is for the regulated 'institutional' bitcoin futures exchanges (N = 2, with **CME** and **Bakkt**); and the other (with N = 8) is for all the most influential instruments from the unregulated bitcoin spot and derivatives exchanges, and the CME. The figures below summarise our results.

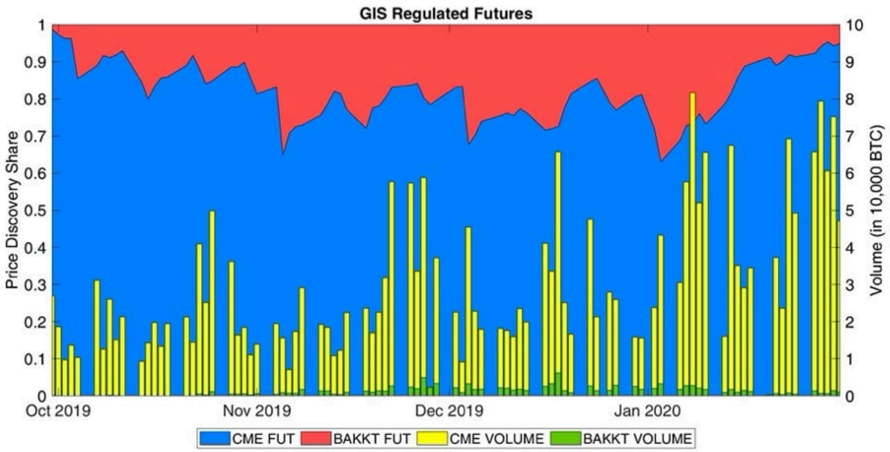


Figure 1: Generalised Information Shares and Volumes: CME and Bakkt Bitcoin Futures

Bakkt, a subsidiary of ICE, eventually launched their (physically-settled) bitcoin futures in September 2019, after considerable media coverage. However, after an initial surge of interest, they have (as yet) failed to capture market share from the CME. In January 2020 the ADV on Bakkt futures was only $12 million, compared with $450 million on CME futures. So, it is not surprising that about 95% of new information – on institutional bitcoin platforms – is led by traders on the CME.

However, the vast majority of off-chain bitcoin trading is on unregulated exchanges. In fact, since its inception in 2017 and until very recently, the BitMEX perpetual swap was by far the strongest leader of bitcoin price discovery[4]. But during the last 12 months the situation has changed. To see this, consider Figure 2, which depicts the GIS for an 8-dimensional VECM. These 8 bitcoin spot and derivatives markets are colour coded and ordered by their final GIS on 31 January 2020.

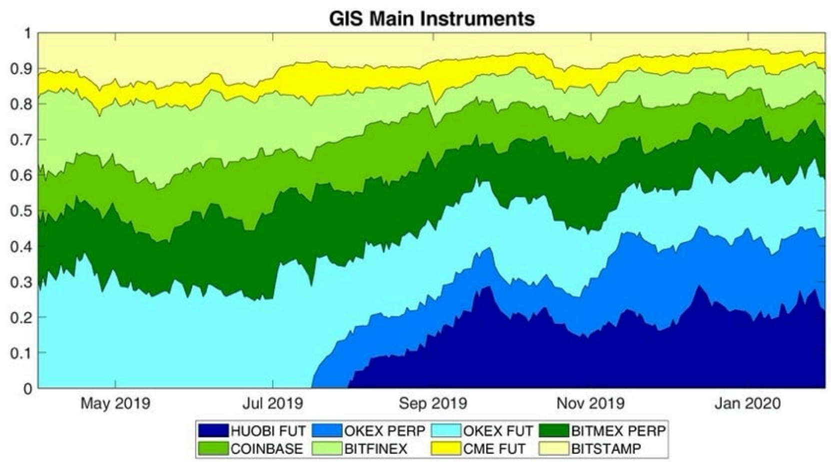


Figure 2: Generalised Information Shares for Bitcoin Perpetual Swaps, Futures and Spot

First, it is evident that the CME futures (shown in yellow) only account for about 6% of total bitcoin price discovery in this system. And even though the Huobi exchange didn't introduce their (quarterly) futures until 2019, this contract has recently become the overall price leader. Its GIS (shown in dark blue in Figure 2) has rapidly grown to take the major share, standing at over 20% on 31 Jan 2020. And its trading volume is second only to the BitMEX perpetual swap (ADV: $2.8 billion on Huobi futures vs $3.2 billion on BitMEX perpetual swap). To put these volume figures in context, for the three spot markets shown in Figure 2 (Coinbase, Bitfinex, and Bitstamp) the ADV in total is only $200 million.

But of all the platforms considered here, OKEx is perhaps the most interesting. By the end of January 2020 their quarterly futures (light blue) and perpetual swap (mid blue) together accounted for over one-third of bitcoin price discovery. Yet, with an ADV of about $1.5 billion – $1.1 billion on the futures and $400m on the perpetual – they have less than one-sixth of the trading volume.

Huobi and OKEx attract better-informed players in bitcoin markets. Located in Singapore and Hong Kong respectively, most institutions are not allowed to trade on these exchanges. Nevertheless, they can look into their order book for high-frequency trading signals. For instance, if you trade on Coinbase (or CME, or Bitstamp, etc.) how many minutes do you have to profit from following a large upward or downward price jump on the OKEx perpetual? How long do you have following a similar move in Huobi futures?

I'll be discussing these results during my talk on Wednesday 13 May. Hope to see you there.

References:

[1] See the CryptoCompare Exchange Review, December 2019

[2] On- and off-boarding the exchange is only captured on a blockchain if it is in cryptocurrency. Transfers in fiat currency are not recorded.

[3] See Alexander C. and M. Dakos (2020) 'A Critical Investigation of Cryptocurrency Data and Analysis' Quantitative Finance. 20:2, 173-188.

[4] See Alexander C., Choi, J., Park, H., and S. Sohn (2020) 'BitMEX Bitcoin Derivatives: Price Discovery, Informational Efficiency and Hedging Effectiveness.' Journal of Futures Markets. 40 (1). pp. 23-43. ISSN 0270-7314.

Carol Alexander is Professor of Finance at the University of Sussex Business School and leader of the Quantitative Finance and Fintech research group. At Sussex she supervises collaborative research on crypto assets and their derivatives with a team of PhD and other research students. For further information see http://www.sussex.ac.uk/profiles/2765

Acknowledgements

We would like to thank the experts featured in this QuantMinds eMagazine for sharing their expertise with us. We look forward to hearing more about what latest research and insights at QuantMinds International in May 2020.

Marcos Lopez de Prado

Professor of Practice at Cornell University, and Co-Founder & CIO, True Positive Technologies

Alexandre Antonov

Chief Analyst, Danske Bank

Jessica James

Managing Director, Senior Quantitative Researcher, Commerzbank AG

Marcos Costa Santos Carreira

PhD Candidate, École Polytechnique

Carol Alexander

Professor of Finance, University of Sussex and Visiting Professor, Peking University HSBC Business School at Oxford

Find more expert insights on the QuantMinds content home >>

Do you want to be involved in a future eMagazine or an event?

Get in touch for speaking opportunities:
Vincent.Beard@informa.com
Get in touch for sponsorship opportunities:
Josef.Lanjri@informa.com
Madeleine.McColgan@informa.com