Perpetual Complexity: an introduction to perpetual future arbitrage mechanics (Part 1)

Introduction

Before any mention of arbitrage is made, it's crucial to understand what a perpetual is. A perpetual is, at its essence, a futures contract that never expires. Perpetuals are some of the most liquid assets in crypto markets and allow much more leverage than an underlying which makes them attractive to opportunists like ourselves for arbitrages. Unlike a traditional futures contract that converges towards the spot when it expires, the perpetual has no reason to stay close to the spot as, well, a deliverable underlying doesn't technically exist. This is why a mechanism called the funding rate is introduced; to ensure the assets remained tied together. The funding rate is absolutely imperative to understand the rest of the article and the proposed strategies. Here's the mechanism: when the perpetual is above the spot price, the funding rate is positive and longs pay shorts, incentivizing people to sell which bring the assets together. Similarly, if the perp is below the spot, the funding rate is negative and shorts pay longs. The funding rate tends to behave in a sticky way and generally follows an autoregressive process, which means, for example, if the funding rate is high now, it tends to be high in the future as well.

Throughout this article, we will focus on exploring two particular subsets of perpetual arbitrage: funding arbitrage and cross-exchange arbitrage. We aim to guide the reader in understanding the mechanics of the strategies as well as the complexities inherent in executing upon these strategies.

Spot-Perpetual Arbitrage

In this article, funding arbitrage mainly refers to two types of arbitrages: spot-perp and perp-perp.

Firstly, spot-perp is a strategy which seeks to profit from the funding rate while remaining delta-neutral (long the spot and short the perp or vice-versa). For example, if the perp is above the spot, this would mean that the funding rate is positive, and under this scenario, we would want to be short the perp and long the underlying. In particular, this is usually the scenario we want to be in. This is because a negative funding rate, albeit profitable with our strategy of long perp and short underlying, will cause the borrowing costs from the spot to eat away most of the edge, which is why market regime does matter a lot in terms of how profitable the strategy is. With this being said, even borrowing costs on the

This strategy is mainly executed on the same exchange because it allows us to be more leveraged, a very important part of this strategy.

As it stands right now, spot-perp sounds quite rudimentary, right? However, this is far from the case: as the reader will come to learn throughout the article, executing upon any perpetual-based arb opportunity is much harder than the ideation phase. Let's get into a couple for spot-perp: the goal with this strategy is to stay in the trade as long as possible. The longer we trade, the more we would profit from the funding rate payments, which is our main source of profit. Although we can still make money from the spread converging, this is not our main focus. So, because we want to be sure that our funding payments are larger than the trading costs, which can be mainly calculated before we enter the trade, we need to model the expected holding time which is our unknown variable. This can be done by smoothing the basis using, for instance, an EMA and then calculating using historical data the mean convergence time in the past.

Although as we mentioned before in this strategy that spread is not the main profit source, when the spread spikes, we could get a good entry and get some extra bps of profit. Empirically, using Bollinger Bands for timing the entries and the exits has been proven to be a very good solution.

Perpetual-Perpetual Arbitrage

Next, perp-perp works in a similar way in which we want to short the perp with the bigger funding rate and long the other, profiting from the spread in their funding rates. We want to be in the trade until the spread between the funding rates closes. One of the main advantages over spot-perp is that we can be more leveraged in this case.

Here we also need to consider the differences in the two perps because for example one difference is the funding payment interval. Some exchanges can have a 1 hour interval and some can have 4 or 8 hours. So it is very important to take it into account when we are trading this strategy by normalizing this spread to make sure we are comparing similar contracts. For example, imagine that we are short a perp with funding rate that has an interval of 8 hours and long a perp that has an interval of 1 hour. This means that until we receive our funding payment from the short leg we must pay 7 payments on the long leg, so if the funding rate on the longer interval changes in the meantime we are exposed to losses. We will expand more on this later during the perpetual arbitrage segment.

When it comes to execution we can either take on both legs, which is easier but also means we pay the spread twice, or we can make on one of the legs which is more efficient. The way this would work is we would market make on one of the legs, and then when we get filled, we would immediately take on the other. Here it's important to take into account the slippage that may occur, especially when the exchange is smaller or the asset is more illiquid and the top of the orderbook is shallow. This is why especially when executing with size we should use VWAP to assess whether we can execute correctly on the taker side.

This strategy also comes with some risks, especially when leveraged. For example, if the basis blows out, then you might get liquidated on the short leg, which would cause significant losses (especially considering the liquidation fees which are very high), or the spread could close to soon in which case we could lose money because of the fees.

Cross-Exchange Perpetual Arbitrage

Alternatively, one could be inclined to explore cross-exchange perpetual arbitrage. Put simply, this arbitrage opportunity is one which involves two perpetual futures with identical underlyings across two (or maybe more) exchanges. Due to the strong similarities in payoffs between the two contracts, the bet is one of relative value; we want and expect these two contracts to converge. You may think this market sounds similar to the aforementioned perp-perp strategy, but there's one key distinction: with perp-perp, the trader is still hunting funding settlements and wishes to stay in the trade for as long as possible to farm these opportunities. On the other hand, cross-exchange perpetual arbitrage is time-agnostic if not impatient, as we wish to enter and close our positions as fast as possible as normalized spreads between both perpetuals revert to a mean value.

When explained in this way, this seems like an easy bit of alpha to capture, no? Simply take opposite sides on both legs, and given the payoff structure, we're going to be in good shape. Unfortunately, as is with all seemingly simple opportunities in crypto, this is not the case. First and foremost, the biggest issue is that these payoff structures are NOT identical. In particular, the construction of these perps across exchanges is not done in a standardized way.

For an example, let's take Hyperliquid and Binance: arguably the largest centralized exchange (CEX) and decentralized exchange (DEX), respectively. Additionally, let's assume an underlying of BTC for simplicity (although actual production of this strategy would likely employ a different cryptocurrency) .

The first issue we encounter is the difference in funding payment intervals; on Hyperliquid, funding payments happen hourly, while Binance has them occur every 8 hours. This mechanically distorts the convergence mechanism of the two, as the funding payment is the catalyst for convergence of the perpetual to spot. Fortunately for opportunists such as ourselves, this alone can be moderately normalized by accounting for accrued interest on

Binance's longer funding interval into the calculation for the current funding rate. This way, both Hyperliquid and Binance's "cleaned" funding rates account for the same mechanics.

However, contributing to this distortion is the difference in calculation of funding rates themselves. Binance and Hyperliquid both use the following formula for funding rates:

$$F = P + clamp(i - P,\ \min F\,, \max F)$$

Where F is the funding rate, P is the average premium index measuring how much the perp deviates from spot, i is a fixed interest rate, and $\min F\,, \max F$ are the maximum and minimum funding rates for a given interval clamping the funding rate. Although both Binance and Hyperliquid use the same clamp values and interest rate across underlyings, the funding cap on Hyperliquid (4% per hour) tends to be higher than Binance's which varies on a case-by-case basis depending on the underlying. Naturally, this distorts the previously-assumed uniform convergence mechanism of the perpetual to the spot between both exchanges, as higher caps means Hyperliquid perps are susceptible to larger spreads in terms of perp-spot. For the aforementioned spot-perp arbitrage, Hyperliquid may be a quite interesting opportunity to explore, as potential funding payments may be larger and hence more attractive, but the increased volatility means borrowing costs and margin may move against you, hence increasing risk to match the relative reward upside. As one can imagine, this becomes a quite complex issue which needs to be normalized for cross-exchange perpetual arbitrage in order to achieve stationarity for future entry-exit rules which will be discussed later on.

Additionally, the construction of the index P varies across both exchanges; Hyperliquid measures P by the following formula:

$$P = \frac{\Delta Impact}{Oracle}$$

Where $\Delta Impact$ is the difference in impact pricing between bids and asks for a sufficiently large notional amount (determined by Hyperliquid on a case-by-case basis with regards to the underlying), and Oracle is simply the price of the underlying as sampled by an oracle every 3 seconds. This makes the premium calculation liquidity-aware, which makes for complications for normalization with an exchange like Binance which uses the following, more simple formula:

$$P = \frac{Mark - Index}{Index}$$

With the $Mark$ price using last traded (or top-of-book) prices and $Index$ prices being calculated as a weighted average across various exchanges for the spot of the underlying. Naturally, this type of formula fails to consider book depth in the same way as the former. For the reader, we hear you; this is really getting in the weeds now and the complexity is piling up, but these are the necessary nuances to understand and normalize for if one wishes to make a truly robust cross-exchange perpetual arbitrage systematic strategy!

Another issue we have is fees: typically, exchanges do not use identical fee structures, meaning that this too must be normalized across perpetuals. Hyperliquid is a DEX which makes the perp experience purely peer-to-peer with more 'generous' maker-taker fees as opposed to CEXs such as Binance. Naturally, this adds a common deterrent for trading smaller convergences, meaning our expected reversions must be large enough in distance to offset transaction costs. Additionally, these fees impact the pricing of the perp itself, as assuming all else equal, Hyperliquid would be priced at a slight premium to Binance due to the implied frictions. Since fee structures tend to vary across exchanges and change (sometimes) materially across volumes, underlyings etc., this is another tricky normalization. If we wanted to create an approximation for the normalization, we could simply add back some of the fees to the Binance perpetual price using fee calculation, but this is not always an easy problem to solve.

Adding to this problem is depeg risk; if any legs have settlements based upon different stablecoins (BUSD, USDC, USDH), a depeg event can seriously hinder profitability due to sudden differences in funding payments in dollar amounts. Assuming these depeg edge scenarios are not considered (a gross simplification), normalizing based upon small differences in stablecoin valuations is advised and fairly straightforward to do. We must take the exchange-native conversion of the stablecoin and account for this rigorously in our normalization process. The reason it's so crucial is because with this strategy, we're looking for bps for edge, and if the funding settlement is based upon different values, this can result in false signals which may lead to losing positions.

Now that we've addressed some of the most glaring issues regarding normalization, let's talk exits and entries. Just like the perp-perp arbitrage opportunity discussed previously, taking on both sides will usually lead to unnecessary fees eating away at PnL. This is why making on the entry leg (and if possible on the exit leg) is seen as the most effective way to optimize.

Now, for entries and exits, we're naturally going to be working with the spreads of both perpetuals. In order to smooth these spreads, it's common to use a simple moving average or an exponentially moving average. After this, we compute the z-score of the spread and trade significant deviations. The reason being that smaller deviations will leave us with negligible returns due to transaction costs eating our profits and a worsened risk-reward profile.

A rule of thumb amongst traders is to trade an absolute z-score of at least two for entries and naturally try to exit as this z-score reverts back to the mean. At this level, some 'touch' is required, as we must account for mean-estimation errors and thus would possibly prefer to exit at an absolute z-score slightly higher than zero. Intuitively, this will mean lower returns, but this may be a necessary trade-off if spreads aren't volatile enough. This is also a good idea due to the diminishing risk-reward profile as the absolute z-score approaches 0. To clarify, for entries, we would prefer to buy on the cheap exchange and sell the expensive one.

Conclusion

As the reader has probably noticed by now, perpetual arbitrage is **no free lunch**. It requires extensive knowledge of statistical properties embedded in market regimes, normalizations amongst other concerns. For spot-perp, opportunities are selective and strongly dependent upon market conditions such as bearish or bullish price action which directly impact the sign of funding rates. Moreover, if negative funding rates arise, our funding payment profit is eaten away by high borrowing costs. On the other hand, both spot-perp arbitrage and perp-perp arbitrage requires the assumption of autoregressiveness of funding rate signs to hold true in order to extract any meaningful excess returns. Additionally, perp-perp arbitrage and cross-exchange perpetual arbitrage require an intensive normalization process on a case-by-case basis. Normalize any of these futures incorrectly, and one of your legs may get liquidated as the market tests your accuracy under stressful intervals. Pair all of these strategies with an intensive fee structure, and you've got yourself a very difficult arbitrage opportunity to harness.

In the future, we wish to use this foundation of perpetual arbitrage mechanics and their inherent risks to create a systematic strategy which addresses the many issues we've discussed throughout the article. This is just the beginning, and more robust explanations of implementation will be revealed soon.

References

[1] Dai, Min et al., "Arbitrage in Perpetual Contracts", 2025

[2] Quant Arb, "Small Trader Alpha #6 – Perpetual Arbitrage", 2024

[3] Quant Arb, "Small Trader Alpha #4 – Funding Arbitrage", 2024

TAGS: Perpetual Futures, Cryptocurrency, Arbitrage, Perpetual Arbitrage, Systematic, Algorithmic Trading

