New Framework for Funding Rates

team@vest.exchange

October 2022

Introductory Notes

What follows below is research carried out by Vest Labs with the primary intent of it being discussed on the labs' forum. This particular session can be found here: https://vest.discourse.group/t/new-framework-for-funding-rates/34

Motivation

We aim to design a perpetual futures exchange for long-tail underlyings. We choose to design an AMM-based exchange rather than a limit orderbook. Because an AMM-based exchange will match buyers (longs) and sellers (shorts) asynchronously, it is possible for an imbalance to exist, exposing the exchange to inventory risk due to implicitly being the counterparty to every trade. Hence, we look towards the current landscape of perpetual futures exchanges as well as literature on (high-frequency) market making to explore how to design a funding rate mechanism that can correct this imbalance.

Current funding rates

Since perpetual futures do not have an expiration date for settlement, [1] originally envisioned funding payments as a mechanism for daily settlement and [2] derives model-free expressions for funding payments using a no-arbitrage argument.

In orderbook-based exchanges, funding payments periodically settle the difference between mark and index price, thereby ensuring mark price to closely track index price. More formally, funding rate F at some time t is defined as

$$F = P + \text{clamp}(I - P, -\delta, \delta)$$

$$P = \frac{(\text{bid} - \text{index})_+ - (\text{index} - \text{ask})_+}{\text{index}}, \ I = \text{quote interest} - \text{base interest}$$

$$\text{clamp}(I - P, -\delta, \delta) = \begin{cases} I - P & \text{if } -\delta < I - P \leq \delta \\ -\delta & \text{if } I - P \leq -\delta \\ \delta & \text{if } I - P > \delta \end{cases}$$

where P represents the premium, I approximates the risk-free rate[1], and the clamp function ensures that the funding rate equals the risk-free rate if P and I are within $\pm\delta$. While the exact implementation may differ, such as specifying the order size that determines impact bid/ask price or a different derivation of the risk-free rate term, most orderbook-based exchanges share this formula[2]. Importantly, most centralized exchanges settle funding payments every 8 hours (e.g. BitMex, Binance) or every 1 hour (e.g. FTX), which concentrates trades around the settlement period.

On the other hand, AMM-based exchange funding rates are typically defined differently. Because longs and shorts are not matched one-to-one, there might exist a long-short imbalance. This imbalance imposes

---

[1]See [3] for an approach to approximate risk-free rates in crypto

[2]See [4] for a comprehensive analysis of funding rates

inventory risk on the exchange because an AMM-based exchange is implicitly the counterparty to all traders. As a result, funding payments can no longer just be a periodic settlement from one side of the contract to the other, but must act additionally as an incentive mechanism for traders (arbitrageurs) to help correct the long-short imbalance by opening positions on the underweight side and receiving funding payments. Hence, funding rates are usually defined to scale with the long-short imbalance $L - S$ where $L, S \geq 0$ denote the total open interest of longs and shorts respectively.

For example, Deri defines funding rate $F = fa(L - S)$ where a determines price impact and f is a function of total liquidity [5] while Kwenta (and Synthetix) defines funding rate $F = \text{clamp}(\frac{L-S}{L+S}, -1, 1) \cdot i_{max}$ where $i_{max} = 10\%/\text{day}$ [6].

Overlay similarly defines funding rate $F = 2k \cdot \frac{|L-S|}{L+S}$ [7], but rather than transferring the funding payments between the traders' margin accounts, settlement is done by decreasing overweight open interest and increasing underweight open interest. As such, assuming the worst case scenario where no arbitrageurs help correct the long-short imbalance, k is calibrated such that with high confidence, the value-at-risk of the imbalance decays to zero at some future time.

Market making literature

The problem of determining a funding rate has parallels to the problem of high-frequency market making. In particular, we examine Avellaneda-Stoikov models [8], which use stochastic control to solve for optimal bid and ask quotes by modelling the arrival of traders on a limit orderbook while accounting for inventory risk. Typically, these models consider a market maker who maximizes her CARA utility. More precisely, the market maker has an objective function of the form

$$\text{maximize} \quad \mathbb{E}[-\exp(-\gamma(X_T + q_T S_T - \psi(q_t)))]$$

where at some time T, X_T is cash, q_T is number of stocks, S_T is stock price, and ψ penalizes inventory, taking forms like $\psi(q_t) = l(|q_t|)$ where l is a liquidity premium for inventory or $\psi(q_t) = \frac{1}{2}\gamma\sigma^2 \int_0^T q_t^2 dt$ where holding an inventory is also penalized over time [9]. However, a limitation of these stochastic control approaches is the over-simplification of market dynamics to make the problem analytically tractable, such as assuming CARA utility, Brownian motion of stock prices, and Poisson process of order arrival rates.

Risk-based funding rates

Although funding rates should be designed to both incentivize arbitrageurs to help correct the long-short imbalance and to compensate the exchange for taking on inventory risk in the temporary absence of arbitrageurs, we will first look at the latter objective. This is because a funding rate mechanism is particularly crucial for any AMM-based exchange. Designs such as Perpetual Protocol v1 that relied on paying out funding from the exchange itself resulted in insolvency in the event of an adverse price movement [10]. At the same time, it is trivial to charge an arbitrarily large funding rate. Note that funding payments should also be a continuous payment from traders' margin accounts to avoid frontrunning risk.

Assuming the worst case scenario, i.e. the absence of arbitrageurs, we set funding rate such that the total funding payment received makes our inventory risk acceptable. We can capture inventory risk with a monetary risk measure and allow the funding payments to decay the inventory risk to zero over time. If the funding rate is less than the risk-free rate, we will simply charge the risk-free rate. More precisely, we require total funding *payment* $\bar{F}_t$ to satisfy

$$\bar{F}_t \cdot \int_s^{s+\tau} e^{-\gamma(s-t)} ds = \rho_{t+\tau}(X) - \Pi_t$$

where $\rho_{t+\tau}$ is a monetary risk measure evaluated over the time horizon from t to $t + \tau$, X is inventory, Π_t is the cumulative funding revenue, and γ is a discount factor. Note that instantaneous funding rate at time

$t + \Delta t$ is $F_{t+\Delta t} = \frac{\bar{F}_t \cdot e^{-\gamma \Delta t}}{P_{t+\Delta t} \cdot \max(L,S)}$ for $\Delta t \in [0, \tau)$. If a new transaction happens at time $t' < t + \tau$, we will recompute $\bar{F}_{t'}$, which defines finding rate for time $[t', t' + \tau)$.

Now, consider a long-heavy market $L > S$. If we simply set X to be the imbalance liability (a short position with open interest $L - S$), then the funding payments provide only the bare minimum for covering the exchange's inventory risk, and there will be no funding payments left for the shorts to receive. To incentivize arbitrageurs who may take on potential market risk to help correct the imbalance, we need to compensate them sufficiently — a risk-based amount is reasonable. So instead, set X to be a short position with open interest L and split the funding payments received pro-rata between the exchange and the shorts. A similar argument holds if the market is short-heavy.

We can choose $\gamma \propto \frac{|L-S|}{L+S}$ so that as the long-short imbalance relative to the total open interest increases, the exchange and arbitrageurs will receive more funding payments, or equivalently receive the total funding payment $(\rho_{t+\tau}(X) - \Pi_t)$ faster. We can choose τ to be relatively small, such as 1 day, to reduce model risk.

Note that one limitation of this risk-based formulation is funding overcharge, despite being in part mitigated by the Π_t term. Intuitively, the overweight side will consistently pay for the worst case scenario but will never receive a refund if that scenario does not happen. One possible solution is to choose ρ to be the expectile value-at-risk so that it also captures the upside from the imbalance liability.

Adaptive funding rates

The above formulation decreases risk by compensating the exchange for the imbalance liability. Risk can alternatively be decreased by correcting the imbalance through adjusting the funding rate to reflect the demand of arbitrageurs. Hence, an alternative formulation is to let the total funding payment consist of two separate parts: a base compensation to the exchange for inventory risk and an additional floating rate that adapts to arbitrage demand.

Setting this floating rate is a trade-off between exploration and exploitation, commonly addressed in reinforcement learning. In fact, various market making literature have explored the application of reinforcement learning for choosing strategies such as setting an optimal bid ask spread [11], given it assumes no knowledge of the market environment and offers a model-free alternative to stochastic control approach literature.

Moreover, several recent papers have explored deep reinforcement learning to combat dimensionality issues. [12] introduces a crypto market-making framework using advanced policy-gradient algorithms and finds favorable results on real limit orderbook data. Deep value-based algorithms like double deep Q-networks, when coupled with adversarial reinforcement learning, have also significantly outperformed to benchmarks like extensions to the Avellaneda-Stoikov model [13].

For our application, we can consider the following: define a floating rate $f = \gamma \cdot \frac{L-S}{L+S}$ similar to many of the existing exchanges, and let $\gamma \in G$ be a variable for some finite set $G = \{\delta, 2\delta, \cdots, (A-1)\delta, A\delta\}$ and $A > 0$. By defining our reward function for choosing γ_t as $r_t(\gamma_t) = (\Pi_{t'} - \Pi_t) - (X_{t'} - X_t)$ where we observe a new transaction at time t', we can apply multiplicative weights algorithm [14] to find optimal $\gamma_{t'}$. With additional constraints on the reward function, we can further extend the learning model to continuous space by applying continuum-armed bandit algorithms [15]. It remains to be shown however if this can be practically implemented in crypto.

References

[1] Robert Shiller. Measuring asset values for cash settlement in derivative markets: hedonic repeated measures indices and perpetual futures. 1993.

[2] Guillermo Angeris, Tarun Chitra, Alex Evans, Matthew Lorig. A primer on perps. 2022.

[3] Lily Francus. Risk frees and currencies. 2022.

[4] UXD Protocol. Comparison of Perpetual Future Funding Rates. 2022.

[5] Deri Protocol. Deri Protocol V3: The Evolution of Capital Efficiency. 2022.

[6] Kwenta. L2 Pereptual Contracts. 2022.

[7] Michael Feldman, Adam Kay, Anantdeep Parihar, Rachel Rybarczyk, and Jonah Lin. Overlay V1 Core. 2022.

[8] Marco Avellaneda and Sasha Stoikov. High-frequency trading in a limit order book. 2008.

[9] Olivier Guéant. Optimal market making. 2017.

[10] Perpetual Protocol. Sunsetting Perp V1. 2022.

[11] Ben Hambley, Renyuan Xu, and Huining Yang. Recent Advances in Reinforcement Learning in Finance. 2022.

[12] Jonathan Sadighian. Deep Reinforcement Learning in Cryptocurrency Market Making. 2019.

[13] Bruno Gašperov, Zvonko Kostanjčar. Market Making With Signals Through Deep Reinforcement Learning. 2021.

[14] Jacob Abernethy and Satyan Kale. Adaptive Market Making via Online Learning. 2013.

[15] Nicolas Della Penna and Mark D. Reid. Bandit Market Makers. 2011.