Perpetual Futures on Vega

W. Gawlikowicz and D. Šiška
witold@vega.xyz and david@vega.xyz

3rd November 2023

Contents

1 Introduction

Vega protocol is a blockchain built from the ground up for trading derivatives [2]. One of the most popular derivatives on centralised exchanges focusing on derivatives of crypto assets are the perpetual futures. The aim of this note is to describe how those could work on Vega.

2 Price tracking in a theoretical perpetual futures market

In the simplest setting there is an underlying asset trading somewhere (on a Vega spot market, on an Ethereum constant function market (CFM) on a centralised exchange...) with price $(S_t)_{t\geq 0}$.

There is a funding interval $\Delta t > 0$ which can be e.g. 8 hours though we will work with years as our time units in which case $\Delta t = \frac{1}{3}\frac{1}{365.25} = \frac{1}{1095.75}$. This gives us a sequence of funding payment dates $t_i := i\Delta t$, $i = 0, 1, \ldots$. There is continuously compounding interest rate r (equivalently a risk-free asset that grows at the constant rate r).

Moreover there is a market where participants bid to enter into a long or short *perpetual futures contract* at a *strike* $(F_t)_{t\geq 0}$. The contract is perpetual and cannot be terminated. However at time t it is possible to go back to the market and trade in the opposite direction (i.e. if long 1 perpetual futures contract then enter a 1 unit short contract and vice-versa) at strike F_t. If at time $t = 0$ two parties agree on strike F_0 then:

1. At $t = 0$ no payments are made[1]; i.e. the contract has current value 0.
2. While the contract is in place the long pays the short, at every t_i, $i = 1, 2, \ldots$ the amount $G_i = \frac{1}{\Delta t}\int_{t_{i-1}}^{t_i}(F_u - S_u)\,du$ i.e. the observed average difference between the futures market strike and the underlying.
3. At termination long gets $F_t - F_0$.

A simple no-arbitrage argument will show that the strike, which from now on we'll refer to as the perpetual futures price, must track the underlying price.

Assuming the funding payments are frequent (and thus taking $\Delta t \to 0$) and summing up the payments we see the present value of the funding payment the long side pays is

$$\int_0^t e^{-ru}(F_u - S_u)\,du\,.$$

We thus get a similar result to [1] on how perpetual future strike and underlying price are related.

Proposition 1. *Assume that the underlying and strike prices are given by some stochastic process* $(S_t)_{t\geq 0}$ *and* $(F_t)_{t\geq 0}$ *adapted to a filtration* $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$.

Assume that there is some risk neutral measure under which the underlying is a martingale. Let us use $\mathbb{E}$ *to denote expectation under this risk-neutral measure and assume that the expected discounted value is integrable in the sense that* $\mathbb{E}\int_0^\infty e^{-ru}|S_u|\,du < \infty$ *and assume that the strike process satisfies the same i.e.* $\mathbb{E}\int_0^\infty e^{-ru}|F_u|\,du < \infty$. *Then either there is arbitrage or the current underlying and perp strike prices satisfy*

$$F_0 = S_0(1 + r)\,.$$

[1] An exchange will expect the parties to deposit margin balance, for now we ignore this

3 Practical considerations

3.1 Marking to market

Rather then waiting until position gets unwound to charge or pay the final cashflow $F_t - F_0$ all position in the market get marked to market with a predefined frequency (up to the nearest block time) whereby each party gets charged or paid the amount $V(F_t - F_{t-\Delta})$, where V is their open volume (negative for shorts), F_t is the new mark price on the perpetual futures market, $F_{t-\Delta}$ is the mark price from the last time a given position has been marked-to-market or modified and $\Delta > 0$ is the length of the interval between mark-to-market updates (or position change).

All positions will be automatically managed (a forced closeout may occur) depending on their available collateral and current margin requirements. For more details see Section 3.3.

3.2 Exchanging funding payments

We have three sequences of times:

1. the underlying price sampling times $(u_k)_{k\in\mathbb{N}}$,
2. the mark price update times $(s_k)_{k\in\mathbb{N}}$ and
3. the funding payment times $(t_k)_{k\in\mathbb{N}}$.

Heuristically the funding rate at time t_k should be

$$R_{t_k} = \frac{\text{mark price TWAP} - \text{index TWAP}}{\text{index TWAP}}$$

So for time t_{k-1} and t_k we want to calculate G_k.

Let $K = \max(j \in \mathbb{N} : u_j \leq t_{k-1})$ and $K^+ = \max(j \in \mathbb{N} : u_j \leq t_k)$. Let $\Delta u_k = u_{k+1} - u_k$ for $k > K$ and $k < K^+ - 1$, let $\Delta u_K = u_{K+1} - t_{k-1}$ and $\Delta u_{K^+} = t_k - u_{K^+}$. Then we calculate index TWAP[2] as:

$$\hat{S}_{t_k} = \frac{1}{t_{K^+} - t_K} \sum_{k=K}^{K^+} S_{u_k} \Delta u_k \;\text{ if }\; K^+ > K \;\text{ and }\; \hat{S}_{t_k} = S_{t_K} \;\text{ otherwise}\,.$$

Similarly we need to get the mark price TWAP. To that end let $L = \max(j \in \mathbb{N} : s_j \leq t_{k-1})$ and $L^+ = \max(j \in \mathbb{N} : s_j \leq t_k)$. Let $\Delta s_k = s_{k+1} - s_k$ for $k > L$ and $k < L^+ - 1$, let $\Delta s_L = s_{L+1} - t_{k-1}$ and $\Delta s_{L^+} = t_k - u_{L^+}$. Then the mark price TWAP applicable in the period, which we call $\hat{F}_{t_k}$, should be:

$$\hat{F}_{t_k} = \frac{1}{s_{L^+} - s_L} \sum_{k=L}^{L^+} F_{s_k} \Delta s_k \;\text{ if }\; L^+ > L \;\text{ and }\; \hat{F}_{t_k} = F_{t_L} \;\text{ otherwise}\,.$$

At time t_k the long should pay the short the amount

$$G_{t_k} = \hat{F}_{t_k} - \hat{S}_{t_k}$$

An alternative way to see the payment is to get it via the funding rate:

$$G_{t_k} = (\hat{F}_{t_k} - \hat{S}_{t_k}) \frac{\hat{S}_{t_k}}{\hat{S}_{t_k}} = R_{t_k} \hat{S}_{t_k}\,, \;\text{ where the funding rate is }\; R_{t_k} = \frac{\hat{F}_{t_k} - \hat{S}_{t_k}}{\hat{S}_{t_k}}\,.$$

Notice that the choice of the denominator is entirely arbitrary and we could equally well have:

$$G_{t_k} = (\hat{F}_{t_k} - \hat{S}_{t_k}) \frac{\hat{F}_{t_k}}{\hat{F}_{t_k}} = \tilde{R}_{t_k} \hat{F}_{t_k}\,, \;\text{ where the funding rate is }\; \tilde{R}_{t_k} = \frac{\hat{F}_{t_k} - \hat{S}_{t_k}}{\hat{F}_{t_k}}\,.$$

[2] time-weighted average price

3.3 Margin considerations

We need to include the estimate of the upcoming funding payment in maintenance margin estimate for the party. Let t_{k-1} be the time of the last funding payment. Let t be current time ($t < t_k$). Calculate G_t as above by replacing t_k with t everywhere. Set

$$m_t^{\text{maint}} = \text{slippage part} + F_t \cdot \text{RF} + \text{margin funding factor} \max(0, G_t).$$

We have margin funding factor $\in [0, 1]$. For more details on how the risk factors RF and slippage part are defined see [3] and [4].

3.4 Ability to replicate properties of existing markets

After reading through even a small portion of documentation on existing perpetual futures markets across different venues one quickly realises that no such thing as a ”canonical perpetual futures contract” exists.

3.4.1 Funding payment calculation

Venues can differ to some extent in various aspects of how $\hat{F}_{t_k}$ or $\hat{S}_{t_k}$ are constructed, but perhaps the most stark difference are observed in how the the funding rate is constructed.

A certain modification that particularly stands out is the introduction of a an interest rate component along with a min and max operator so that the funding rate takes form:

$$R_{t_k} = \frac{\hat{F}_{t_k} - \hat{S}_{t_k}}{\hat{S}_{t_k}} + \min\left(a, \max\left(b, \frac{(1 + (t_k - t_{k-1})r)\hat{S}_{t_k} - \hat{F}_{t_k}}{\hat{S}_{t_k}}\right)\right),$$

where a and b are constants such that $a \geq b$ and r is the interest component which may be constant or time-dependent.

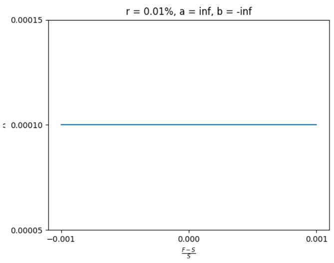

Figure 1:
$r = 0.01\%, a = -b = \infty$.

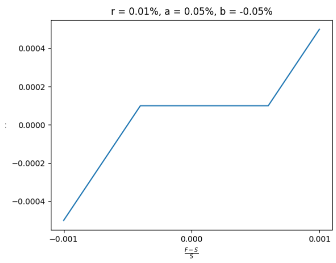

Figure 2:
$r = 0.01\%, a = -b = 0.05\%$.

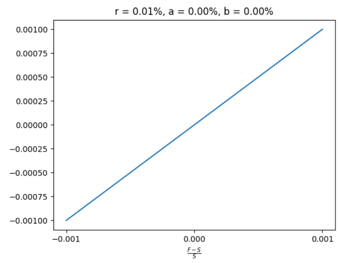

Figure 3:
$r = 0.01\%, a = -b = 0$.

Since Vega Protocol strives to offer adequate flexibility to allow community to shape the markets according to its needs we choose to allow reproducing these feature by prescribing the following form for the implementation of the funding payment component described in section 3.2. That modified funding payment $\tilde{G}_{t_k}$ is then:

$$\tilde{G}_{t_k} = \hat{F}_{t_k} - \hat{S}_{t_k} + \min\left(a\hat{S}_{t_k}, \max\left(b\hat{S}_{t_k}, e^{r_c(t_k - t_{k-1})}\hat{S}_{t_k} - \hat{F}_{t_k}\right)\right),$$

where r_c is the continuous rate of return set for the market (can be updated via governance) and $[t_k, t_{k-1})$ is the time interval in years for which the funding payment is being calculated.

However as the team has not identified any specific use case for the these additional parameters we recommend initially setting $a = b = r_c = 0$, unless setting them to different values solves a specific issue identified by the community.

Some of the properties that can be observed in various venues, but which we deliberately choose not to reproduce are discussed in the remaining subsections.

3.4.2 Tolerance region

There are venues which use the notion of impact bid and ask price to specify the region within which the differences between F_{t_k} and S_{t_k} are considered negligible and thus not impacting the funding payment. The impact bid/ask price is the average fill price of an order of a fixed notional (called impact notional) at the buy/sell side of the order book. The basic funding payment discussed in section 3.2 would then take the form:

$$\hat{G}_{t_k} = (\hat{F}_{t_k} - \hat{S}_{t_k})\mathbb{1}_{(\hat{F}_{t_k} - \hat{S}_{t_k}) \notin A},$$

where

$$\mathbb{1}_{x \notin A} = \begin{cases} 0 & \text{if } x \in A, \\ 1 & \text{otherwise}, \end{cases}$$

and A is some interval, say $A = [(1-c)F_{t_k}, (1+c)F_{t_k}]$, where c is a constant.

However, we choose not to employ this mechanism in our implementation of the funding rate calculation.

3.4.3 Non-homogeneous time weighting

Certain venues modify the time-weighted average price used for funding rate calculation to assign different importance to observations depending on their relative time from the funding payment, however we choose not to employ such a mechanism in our implementation of the funding rate calculation.

3.5 Funding payment and spot sampling frequency

In this section we will briefly consider the impact of funding frequency and spot sampling frequency on the funding payments observed in a market.

We will use the two curves displayed below alongs with 7 day funding frequency, spot curve sampled every 1 hour and perp curve sampled every 5 minutes as our baseline setup.

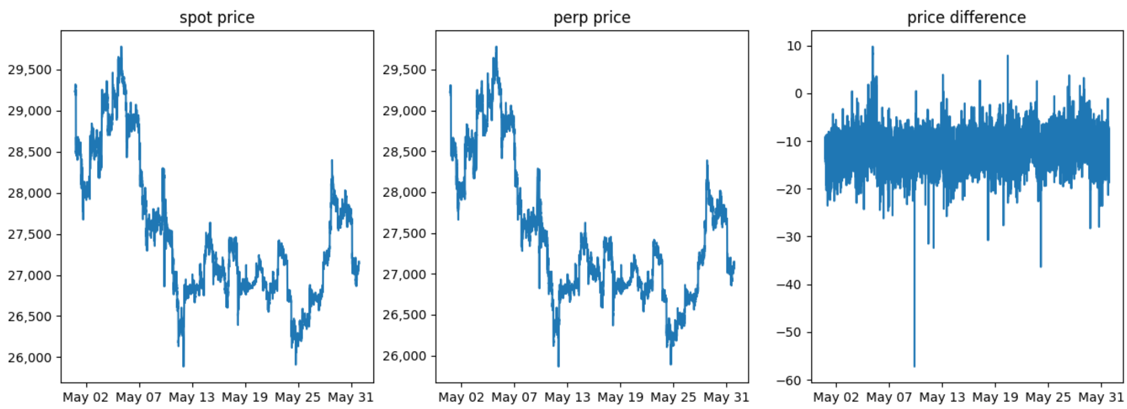

Figure 4: Example BTC/USD spot and perpetual features prices for May 2023.

The resulting funding payments and rate of return for the perpetual contracts of size 1 held throughout such market are displayed in figure 5.

The impact of varying the funding payment frequency is summarised in table 1.

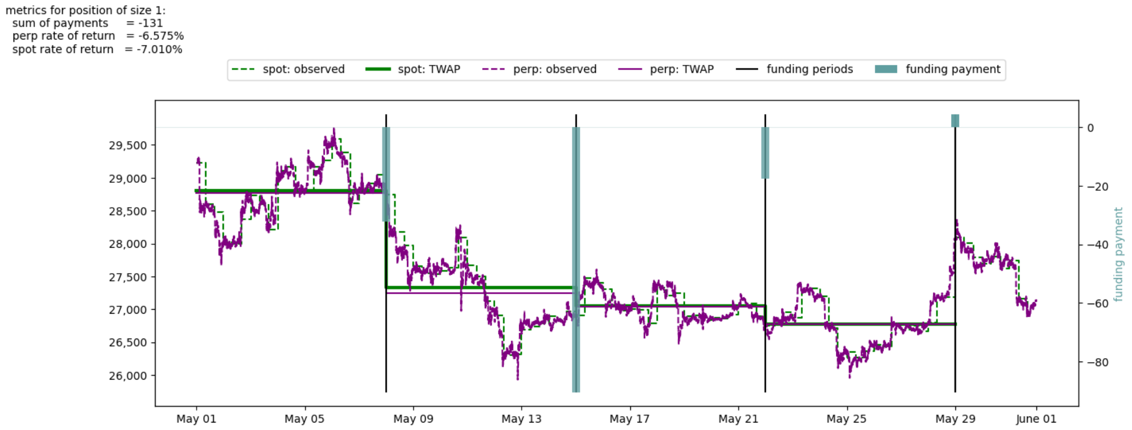

Figure 5: Resulting funding payments for the baseline case. Interactive version can be found at: https:/github.com/vegaprotocol/research.

Table 1: Impact of varying funding frequency

frequency	perp rate of return	sum of funding payments
7d	-0.065605	-1.36E+02
1d	-0.060621	-2.81E+02
12h	-0.006304	-1.87E+03
6h	0.060692	-3.83E+03
3h	0.185106	-7.46E+03
1h	0.736055	-2.36E+04
30min	1.464369	-4.48E+04
10min	4.332158	-1.29E+05
5min	8.684744	-2.56E+05

The impact of varying the spot sampling frequency is summarised in table 2.

Table 2: Impact of varying spot sampling frequency

frequency	perp rate of return	sum of funding payments
7d	0.076308	-4.28E+03
1d	-0.057820	-3.63E+02
12h	-0.063995	-1.83E+02
6h	-0.065458	-1.40E+02
3h	-0.067809	-7.14E+01
1h	-0.068639	-4.71E+01
30min	-0.068613	-4.79E+01
10min	-0.068650	-4.68E+01
5min	-0.068625	-4.76E+01

A Proofs

Proof of Proposition 1. □

Let us start by noting that our assumptions that $\mathbb{E}\int_0^\infty e^{-ru}|S_u|\,du < \infty$ and $\mathbb{E}\int_0^\infty e^{-ru}|F_u|\,du <$

∞ imply that

$$\left|\mathbb{E}\int_0^\infty e^{-rt}(F_t - S_t)\,dt\right| \le \mathbb{E}\int_0^\infty e^{-rt}\,|F_t - S_t|\,\,dt \le \mathbb{E}\int_0^\infty e^{-rt}|F_t| + |-S_t|\,dt < \infty\,.$$

Since the parties make no payment upon entering the contract the value at time 0 must be 0 and thus we must have for every $t \ge 0$ that

$$0 = \mathbb{E}\left[e^{-rt}(F_t - F_0) - \int_0^t e^{-ru}(F_u - S_u)\,du\right]\,.$$

As the discounted underlying is a martingale we must also have, for every $t \ge 0$, that

$$0 = \mathbb{E}[e^{-rt}S_t - S_0]\,.$$

Let us write $\psi_t = \mathbb{E}[e^{-rt}(F_t - S_t)]$. Then for every $t \ge 0$ we have

$$\psi_t = F_0(e^{-rt} - 1) + \psi_0 + \int_0^t \psi_u\,du\,.$$

We can solve this integral equation the same way as we would solve an ODE with an integrating factor (i.e. calculate $d(e^{-t}\psi_t)$ and integrate). This takes us to

$$e^{-t}\psi_t = \psi_0 - rF_0\int_0^t e^{-u}e^{-ru}\,du\,.$$

Evaluating the integral we get

$$e^{-t}\psi_t = \psi_0 - \frac{rF_0}{1+r}(e^{-(1+r)t} - 1)$$

and so

$$\psi_t = e^t\psi_0 - \frac{rF_0}{1+r}(e^{-rt} - e^t)\,.$$

This is

$$\mathbb{E}[e^{-rt}(F_t - S_t)] = e^t(F_0 - S_0) - \frac{rF_0}{1+r}(e^{-rt} - e^t) = e^t\left(F_0 - S_0 - \frac{rF_0}{1+r}\right) - e^{-rt}\frac{rF_0}{1+r}\,.$$

Let $\gamma = F_0 - S_0 - \frac{rF_0}{1+r}$. We see that

$$\mathbb{E}\int_0^\infty e^{-rt}(F_t - S_t)\,dt = \gamma\int_0^\infty e^t\,dt - \frac{rF_0}{1+r}\int_0^\infty e^{-rt}\,dt = \gamma\int_0^\infty e^t\,dt - \frac{F_0}{1+r}\,.$$

Thus

$$\mathbb{E}\int_0^\infty e^{-rt}(F_t - S_t)\,dt = \begin{cases} +\infty & \text{if } \gamma > 0\,, \\ -\dfrac{F_0}{1+r} & \text{if } \gamma = 0\,, \\ -\infty & \text{if } \gamma < 0\,. \end{cases}$$

Note that by assumption we have

$$\left|\mathbb{E}\int_0^\infty e^{-rt}(F_t - S_t)\,dt\right| \le \mathbb{E}\int_0^\infty e^{-rt}|F_t - S_t|\,dt \le \mathbb{E}\int_0^\infty e^{-rt}|F_t|\,dt + \mathbb{E}\int_0^\infty e^{-rt}|S_t|\,dt < \infty\,.$$

But $\left|\mathbb{E}\int_0^\infty e^{-rt}(F_t - S_t)\,dt\right| < \infty$ means that $\gamma = 0$ which means that

$$F_0 = S_0(1+r)\,.$$

References

[1] S. He and A. Manela and O. Ross and V. von Wachter. Fundamentals of Perpetual Futures. *arXiv*:2212.06888, 2023.

[2] G. Danezis, D. Hrycyszyn, B. Mannerings, T. Rudolph and D. Šiška. Vega Protocol: A liquidity incentivising trading protocol for smart financial products. *Vega research paper*, 2018.

[3] D. Šiška. Margins and Credit Risk on Vega. *Vega research paper*, 2019.

[4] Vega protocol specifications. Margin calculator. *https://github.com/vegaprotocol/specs/blob/master/protocol/0019-MCAL-margin_calculator.md*.

