Funding Payments Crisis-Proofed Bitcoin's Perpetual Futures

Preliminary and incomplete

Will Gornall* Martin Rinaldi* Yizhou Xiao†

November 28, 2024

Abstract

We study futures contract design using the volatile cryptocurrency market as a laboratory. In this market, order flow frequently overwhelms arbitrage capital and pushes futures prices above or below their underlying assets. Perpetual futures emerged as a response to this. These contracts tightly track their underlying due to small, frequent payments. We show that these contracts reduced noise trader risk, dominated trading, and improved liquidity; and rationalize those results using a tractable model. We argue that these contracts offer potential financial stability benefits because they improve crisis liquidity and reduce the drawdowns of common arbitrage strategies by more than half.

*University of British Columbia, Sauder School of Business.

†The Chinese University of Hong Kong (CUHK) Business School.

We are grateful to the faculty at Carnegie Mellon University, CUHK, and UBC and Yoontae Jeon for their valuable insights, helpful comments, and suggestions. We would also like to acknowledge the support of Aquanow and the MITACS Accelerate program, which provided funding for this research.

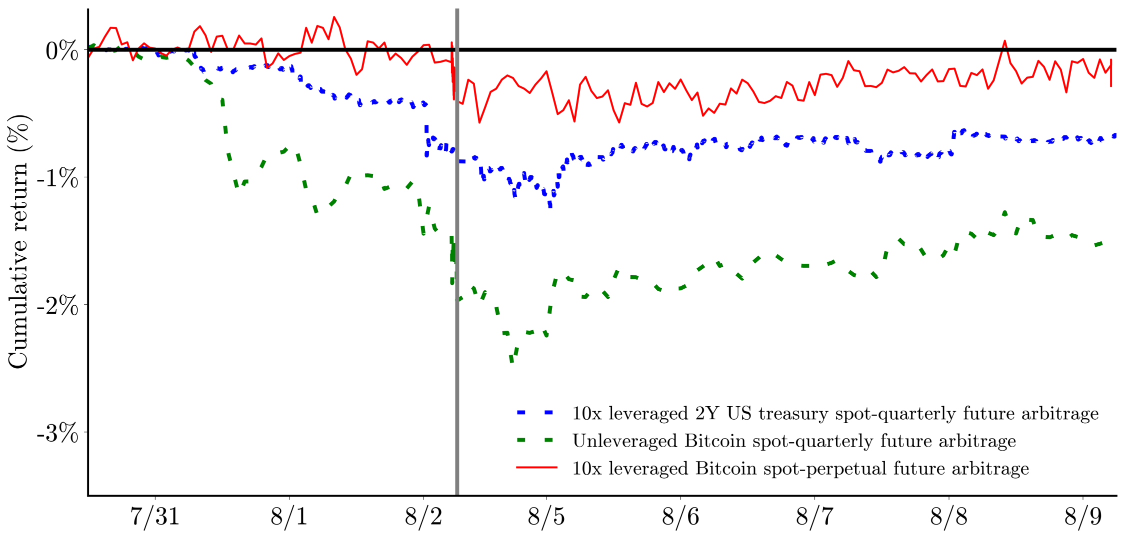
1. Introduction

Some of the most well-known arbitrage strategies are spread trades that pair a long position in an underpriced spot asset with a short position in its overpriced future (or vice versa). These arbitrage strategies are profitable in normal times, but suffer occasional large drawdowns when the deviations that the arbitrageurs are betting against grow instead of shrink. Arbitrage spread trades yielding positive returns with interim drawdowns have been documented in currencies (Du and Tepper, 2018), equity Binsbergen et al. (2012), corporate credit (Bai and Collin-Dufresne, 2019), and even U.S. treasury bonds (Barth et al., 2021).

Figure 1 illustrates the returns to several arbitrage spread trades during a recent global risk-off event. This event was triggered by a surprise Bank of Japan rate hike on July 31, 2024, and a weak US jobs report on August 2. Although far from a crisis, these events tightened the market for arbitrage capital and widened the futures basis even in markets unrelated to Japan.

We first show the return to a treasury cash-futures basis trade, where a trader sells a December 2-year treasury future and hedges that risk by buying a two-year treasury using money borrowed until December at LIBOR, with these trades conducted at 10X leverage. The portfolio is constructed following Glicoes et al. (2024) using data from Reuters (via Refinitiv) with the exposure of each leg of the trade set to 10x the portfolio size, consistent with specialized intermediaries conducting this treasury futures arbitrage with up to 50x leverage (Avalos and Sushko, 2023). As the figure shows, this trade loses about 1% of its value, or 0.1% of its position size. That is a small shock in the absolute sense, however, the underlying market movements were also small. The dramatic futures basis expansion during the COVID-19 crisis and the 2008 financial crisis led to much larger losses on these trades (Siriwardane et al., 2022). Moreover, a loss of 0.1% of gross exposure may be a large shock for a 50x leveraged fund.

Figure 1: Returns to basis arbitrage trades during instability.
This figure plots the cumulative returns to different basis arbitrage strategies around the August 2024 stock market retraction: 1) a 10x leveraged treasury cash-futures basis trade following Glicoes et al. (2024) that uses a December future to hedge 2-year treasuries financed at LIBOR; 2) an unlevered crypto basis trade that uses a Binance December BTC-USDT future to hedge BTC borrowed from lending platform AAVE; and 3) a 10x leveraged crypto basis trade that uses a Binance BTC-USDT perpetual future to hedge BTC borrowed from lending platform AAVE. We set the initial notional exposure of each leg of the trade equal to the portfolio size for the unleveraged trade and equal to 10x portfolio size for the 10x leveraged trades. The verticle line marks the weekend, which is omitted.



The second trade we show is a basis trade on the Bitcoin (BTC) cryptocurrency using quarterly BTC futures and no leverage. Here, the trader buys a December BTC future on Binance and hedges that risk by short-selling BTC using the AAVE cryptocurrency lending market. Given the risk in crypto assets, we assume that this trade is conducted on an unlevered basis. Even with one-tenth the leverage of the treasury basis trade, the BTC basis trade has much larger losses. The BTC basis arbitrage showing much larger draw losses than the treasury basis arbitrage is unsurprising given the volatility of BTC prices and the quality of the treasury market.

Finally, Figure 1 constructs a 10x leveraged BTC futures basis trade using so-called "perpetual futures" instead of quarterly futures. Even at 10x leverage, this perpetual future basis trade has a far less severe drawdown than the unlevered BTC basis trade conducted using the December quarterly future. In fact, the 10x leveraged perpetual future basis trade on BTC has a less severe drawdown than our analogous 10x leveraged December future basis trade on U.S. treasury notes: a stunning result given that two-year note prices fluctuated just 2% during this window, whereas spot BTC prices fell more than 25% from Friday highs to Monday lows and a reported 300,000 crypto traders accounting for more than $1 billion were liquidated.[1] In this paper, we analyze these perpetual futures contracts empirically and theoretically, arguing that they improve market quality, tighten limits to arbitrage, and help the functioning of futures markets during crises.

The payoff of perpetual futures is similar to holding a portfolio of very short-term (often 8-hour) futures that are continually rolled over. Specifically, perpetual futures feature small, frequent payments between the long side and the short side of the contract. The direction of these so-called funding payments pushes the futures price toward its underlying: longs pay shorts when the future price is too high and vice versa. For example, every eight hours, the long perpetual futures holders on BitMEX pay the shorts an amount equal to the weighted average of the difference between the futures price and the spot price over that eight-hour window. These funding payments keep the futures prices close to spot prices even in severe downturns and so the actual payments are small: despite the large spot price deviations surrounding Figure 1, the funding payments ranged from -0.004% to 0.01% paid every eight hours.

We refer to the risk of a futures price diverging from its spot price as basis risk. Because perpetual futures more closely track their underlyings, they have less basis risk than quarterly futures. This basis risk is unimportant for traditional markets in normal times, but the basis fluctuations seen in the 2008 financial crisis or COVID-19 crisis show futures basis risk can matter when limits to arbitrage bind.

Studying limits to arbitrage during crises using traditional financial markets is challeng-

[1] See "Over $1B wiped out in crypto liquidations as global markets suffer" *Cointelegraph*.

ing because they are rare and marked by government intervention (and the expectation of government action). Cryptocurrencies offer an ideal laboratory to study financial crises for several reasons. First, cryptocurrency provides an abundance of crises: while the broad US market has recorded four 10%+ daily declines over the past century, the cryptocurrency market has seen *fifty* such declines in the past *decade*.[2] Second, cryptocurrency lacks any form of policy intervention, allowing us to study how market structure evolves without the prospect of taxpayer-funded bailouts. Third, the unregulated nature and low entry costs of cryptocurrency venues have led to the emergence of a diverse array of markets, structures, and contracts; with more than one hundred venues offering at least one variety of BTC future, each with subtly different specifications.

We use cryptocurrency to establish several empirical results about perpetual futures. First, we document the rise of these contracts. Pioneered on BitMEX in May 2016, these contracts were the end result of a progression from quarterly, to monthly, to weekly and even daily futures expirations. Perpetual futures quickly dominated BitMEX before spreading to almost all other exchanges. They have outcompeting standard futures and nowadays account for almost all cryptocurrency futures trading outside of the US. There are a variety of potential explanations for the popularity of these contracts. We argue that the most natural driver of their popularity was market participant's dislike of basis.

Second, we document that these perpetual futures contracts do indeed reduce basis and, in particular, basis risk. When an asset's price jumps, futures on that asset tend to show even large moves because the futures basis moves in the same direction. For the BTC futures we look at, this basis effect means quarterly futures move 8-10% more than the spot price when shocks occur. Perrpetual futures have less than half as much basis risk and exagerate price movements by about 3%. Basis risk matters relatively more for arbitrage spread trades that hedge a spot position with a futures position, as we illustraded in Figure 1. Perpetual futures

[2] Author's calculations using daily Coinbase data.

make these trades far safer, with the peak losses of spread trades constructed using perpetual futures being less than half the losses of analogous trades constructed using quarterly futures.

Third, we show that perpetual futures are more liquid than quarterly futures. The differences are quantitatively large: across Kraken and BitMEX, and using regressions and matching, we see perpetual futures have spreads that are 53% to 71% lower. Perpetual futures' superior liquidity appears to be even larger during times of market volatility. Although both types of futures become less liquid around large market movements, quarterly futures spreads increase relatively more.

We build a tractable theoretical model that links these empirical results. Our model has a futures contract traded by market makers, liquidity traders, and long-term hedgers. Liquid traders and market makers are risk-averse to mark-to-market trading losses, while long-term hedgers seek to minimize the risk of a timed hedging exposure. Time-series variation in the liquidity traders' demand creates volatility of the basis of the future. We vary the form of the futures contract to evaluate standard futures, perpetual futures, and hybrid instruments that feature both payments at expiration and funding payments.

Under this model, standard futures with short maturities and perpetual futures with large funding payments both track the spot price closely. Reducing basis risk reduces the risk of large mark-to-market losses, which leads to more trade and higher welfare for the market maker and liquidity trader.

Hedgers with randomly occurring long-term hedging needs also prefer perpetual futures, even if they do not care about the short-term. First, perpetual futures reduce their costs because reducing basis risk means market makers require less compensation for offsetting hedger trades. Second, perpetual futures can reduce the overall risk hedgers face, provided hedger cash flows are not disproportionally at the maturities of standard futures. To illustrate, a standard quarterly future is a perfect hedge for a claim at the end of the quarter, but claims that are not at quarter ends create basis risk when their futures are sold prior to expiration or rolled over. This basis risk is greater than the risk of funding payments.

By pegging the futures price to the spot price, these contracts transfer basis risk from the arbitrageur to other market participants. This is undesirable for some users: a utility may want a December natural gas future to hedge a December need because the spot price of hard-to-store natural gas offers a poor hedge. However, for financial futures, the vast majority of trading happens at single active expiration, as Appendix Figure A1 shows. This suggests that, unlike many commodity futures traders, the traders of financial asset futures care about broad market exposure and not a specific delivery time. A mutual fund buying a treasury future may not care about specific delivery and simply want cheap interest exposure with low collateral requirements. These traders choose futures over spot markets not because they offer different payoffs, but because spot exposure would entail negotiating over-the-counter treasury purchases and setting up a bilateral collateral agreement. Because of this, pegging the futures price to the spot price using funding rates seems most suitable for financial assets.

Our paper links several lines of prior work. First, there are two papers discussing perpetual futures in traditional financial markets. Gehr Jr (1988) documents that these originally shored-lived contracts were used by the Chinese Gold and Silver Exchange of Hong Kong and Shiller (1993) proposes offering these futures on illiquid assets as a way to better match hedger needs and encourage the production of long-term information.

After these two early papers, the literature was largely silent on perpetual contracts for two decades until the rise of perpetual futures in cryptocurrency. In the cryptocurrency context, Christin et al. (2022) articulate an arbitrage strategy using perpetual futures and link its outsized returns to speculator demand. Schmeling et al. (2023) similarly describe basis arbitrage ascribing its cause to outsized retail demand and a scarcity of arbitrage capital. He et al. (2022) describe the mechanics of perpetual futures and an alternative speculative strategy to exploit demand imbalance. Alexander et al. (2021) and Deng et al. (2020) construct optimal hedging strategies for cryptocurrency futures.

Second, we link to the literature on limits to arbitrage. The noise trader risk channel we focus on builds on work by De Long et al. (1990) and Foucault et al. (2011). We build on Shleifer and Vishny (1997), extended by Brunnermeier and Pedersen (2009), and applied by Acharya et al. (2013) to commodity futures markets under a relative scarcity of arbitrage capital. Our focus on futures tenor and basis risk leverages Rutledge (1976) and Castelino and Francis (1982) who report that basis risk increases with time to maturity. Our paper is perhaps closest to Hazelkorn et al. (2020), who model and explore the basis of traditional finance contracts and link it to dealer financing costs, and Figlewski (1984), who report on equity index basis and link it to limits to arbitrage.

Third, our paper builds on the contractual innovation and contractual evolution literature, see for example Tufano (2003) or Miller (1986). We argue that these contracts evolved in the crisis-prone world of crypto to complete the market (Duffie and Rahi, 1995) and meet market participants' liquidity needs (Shen et al., 2014). These contracts could improve welfare by lowering hedger costs and reducing basis risk, albeit their effect on cryptocurrency trader welfare is unclear (e.g., Simsek (2013)).

2. Data and measures

2.1. Cryptocurrency trade and orderbook data

We compile granular data from BitMEX, Deribit, and Kraken using a combination of API requests and bulk CSV exports. These were three of the earliest, credible cryptocurrency futures exchanges and the first three exchanges to offer cryptocurrency perpetual futures. Our data span from these three platforms inceptions to a cutoff date of December 31, 2023.

Appendix Figure A4 shows the product offerings of cryptocurrency exchanges offering futures prior to 2019. Although the vast majority of surviving exchanges have adopted perpetual futures, we focus on the first three adopters because we are interested in the benefits of switching to new contracts and when later adopters switched perpetual futures were essentially the standard contract. In addition to being early adopters of perpetual futures, the three exchanges we focus on are the earliest cryptocurrency futures exchanges that did not fail and have not been accused of faking trading volume data by prior work.[3] In particular, the data provided by the other two early adopters (HTX and OKX) does not appear to be reliable.

BitMEX was the first cryptocurrency exchange to offer cryptocurrency perpetual futures. It provides both trade and top-of-the-book data for essentially its entire existence starting November 11, 2014. Several of our microstructure measures rely on a reference price.

Deribit is a cryptocurrency exchange that specializes in options trading. It was the second exchange to adopt perpetual futures. Deribit provides data from its inception in February 2017 onward, including trade and mark-to-market pricing.

Kraken (then CryptoFacilities) was the third perpetual futures adopter and the first regulated exchange to do so. After an acquisition by regulated, US-based cryptocurrency company Kraken, it changed its name to Kraken Futures and so we refer to it as Kraken for simplicity. Kraken also provides data for essentially its entire existence, including trade and mark-to-market pricing data from January 1, 2018.

2.2. Other cryptocurrency data

Spot trading metrics are retrieved from Coin Metrics, a cryptocurrency data provider. We set the spot price equal to the median of the prices from Coinbase, Bitfinex, and Bitstamp, three of the oldest surviving cryptocurrency exchanges and popular exchanges throughout our period of study. We use the median of the last-traded price across these three exchanges,

[3] We consider reported wash trading or fake volume by Chen et al. (2022), Cong et al. (2022), Bitwise's 2019 SEC report, and the Blockchain Transparency Institute. Fake volume and other data manipulation are prevalent in cryptocurrency because exchanges typically combine the roles of exchanges, custodians, and brokers. Combining these roles not only led to numerous frauds (e.g., FTX, Thodex, Quadriga), it allows exchanges to manipulate trade data.

which allows us to adjust for cross-exchange pricing anomalies caused by hacks or panics and aggregate this to a coin-by-second panel. The prices on these three venues closely align, unlike the pricing of venues in countries with capital controls (Choi et al., 2022).

Aggregate volume and product offering data are retrieved from Coin Metrics, a cryptocurrency data provider. These data cover 20 futures venues, which together account for the vast majority of the cryptocurrency futures market. The data includes exchanges that appear to manipulate trading and volume measures, which we exclude from our main analysis.

2.3. Measures of market quality

We use several standard measures to assess the quality of futures markets. We focus on the dislocation of futures prices from spot prices (the futures basis) and various measures of transaction costs. Consider a futures contract traded at time t at average price F_t whose underlying is a spot asset with price S_t. The futures basis, B_t, is the extent to which the price of that futures contract i exceeds the spot price of its underlying asset:

$$B_t = \frac{F_t}{S_t} - 1. \tag{1}$$

We aggregate that to the day level by averaging across the last trade of each minute for minutes that contain trades.

We use a variety of standard measures of transaction cost. We have top-of-the-book data for BitMEX and so define the quoted spread at time t as

$$QuotedSpread_t = \frac{F_t^{ASK} - F_t^{BID}}{F_t^{MID}}, \tag{2}$$

where F_t^{ASK} is the best ask quote at time s, F_t^{BID} is the best bid quote, and F_t^{MID} is their midpoint. For day d, we report the average of the quoted spread across all minute ends.

The Roll (1984) estimator of the spread is based on the square root of the absolute value of the serial covariance of price changes, relying on the premise that the distribution of price changes is stationary over short time intervals. For day d, we define this as

$$RollSpread_d = 2\sqrt{\text{AUTOCOVAR}_{t \in N_d}\left[F_t/F_{t-1}\right]}, \tag{3}$$

where $\text{AUTOCOVAR}_{t \in N_d}$ is the autocovariance calculated over the trades on day d. We discard this measure for symbol-days with fewer than 10 trades.

The effective spread estimates the spread based on executed prices versus a reference price as follows:

$$EffectiveSpread_t = 2 \times \begin{cases} F_t/F_{t-}^{REF} - 1 & \text{if trade } j \text{ is a buyer-initiated} \\ 1 - F_t/F_{t-}^{REF} & \text{if trade } j \text{ is seller-initiated,} \end{cases} \tag{4}$$

where F_{t-}^{REF} is a reference price for the future immediately before the trade. For BitMEX, we set this reference price to the mid-point in the second prior to the trade. For Kraken and Deribit, we set it to the mark-to-market price provided by the exchanges, which is a smoothed version of F_t^{MID} designed to resist manipulation.[4] For Deribit, we lack the initiator of the order so assume that an order is buyer (seller)-initiated if the price is above (below) the mark price. Instead of taking the simple daily average, we set the effective spread for day d equal to half the average of $EffectiveSpread_t$ for buyer-initiated trades plus half the average for seller-initiated trades. We adopt this to reduce any potential bias created by unusually large numbers of buys or sells on a particular day, in practice this has little impact. Finally, we discard this measure for symbol-days with fewer than 10 trades in either mean.

[4] For both Kraken and Deribit, this is set to $S_t + EWM_{30}[F_s^{IMPACTMID} - S_s]$, where EWM_{30} is a 30-second exponential moving average and $F_s^{IMPACTMID}$ is equal to the mid-point between the average price of buying a certain quantity of futures and selling the same quantity and S_s is set based on the exchange's version of the spot reference price. These specifications are provided online by Kraken and Deribit.

3. Empirical results

3.1. Structure and dominance of perpetual futures

Speculators eager for leverage have made cryptocurrency futures contracts extremely popular. While traditional futures contracts have cash payments or physical delivery at maturity (which we refer to as standard futures), cryptocurrency futures use a so-called perpetual structure. The perpetual futures structure was originally adopted by the Chinese Gold and Silver Exchange of Hong Kong, as reported by Gehr Jr (1988), however, it did not take off elsewhere prior to its emergence in cryptocurrency trading.

Rather than using a cash flow or delivery at expiration to push futures prices towards their underlying, these contracts use so-called funding rate payments. These funding rate payments mean perpetual futures have a payoff that resembles extremely short-maturity standard futures that are continually rolled over at the spot price. An idealized version of the perpetual future has the long position pay the short position continuously at rate

$$FundingPayment_t = \lambda(F_t - S_t) = \lambda B_t S_t, \tag{5}$$

where λ is a scaling parameter, F_t is the time-t value of the future, S_t is the time-t price of its spot underlying, and B_t is the basis. When the futures price is above the spot price, the long futures holders pay the short holders, and the expectation of these payments pushes the futures price down. Conversely, payments from short holders to long holders push the futures price up toward the spot price. λ is typically set so that the full difference between the futures price and the spot price is paid from longs to shorts every eight hours.

In practice, each exchange uses a venue-specific variation of Equation (5). Beyond varying λ and the precise calculation of S_t and F_t, exchanges frequently set these payments to occur every eight hours based on the previous eight hours of basis, add an additional interest-rate transfer from longs to shorts, limit the allowed payment to combat manipulation, and shrink the payment toward zero to prevent small payments.[5]

This structure allows for strong arbitrage pressure without requiring a short maturity. For example, suppose a futures contract is trading 2% above its underlying. For a standard future, an arbitrager can earn 2% of the spot price by buying the underlying, shorting the future, and unwinding that trade at the future's maturity (ignoring transaction costs). That may be a sufficient incentive during normal times, but it may be too weak if there is scarce arbitrage capital and the future's maturity is months away. The only way to increase the arbitrager's return is to shorten the maturity.

Adjusting the λ parameter of perpetual futures allows us to dramatically strengthen that convergence pressure without requiring a short maturity. For example, most perpetual futures formulas translate a 2% price deviation into a return of at least 2% per day. The prospect of these payments means the futures price closely tracks its underlying, and so the actual payments made are typically small. For example, Figure 1 corresponded to a 25% drop in the price of the underlying, yet because the futures price tracked the underlying so closely, the thrice-daily funding payments peaked at just 0.01% of the notional value.

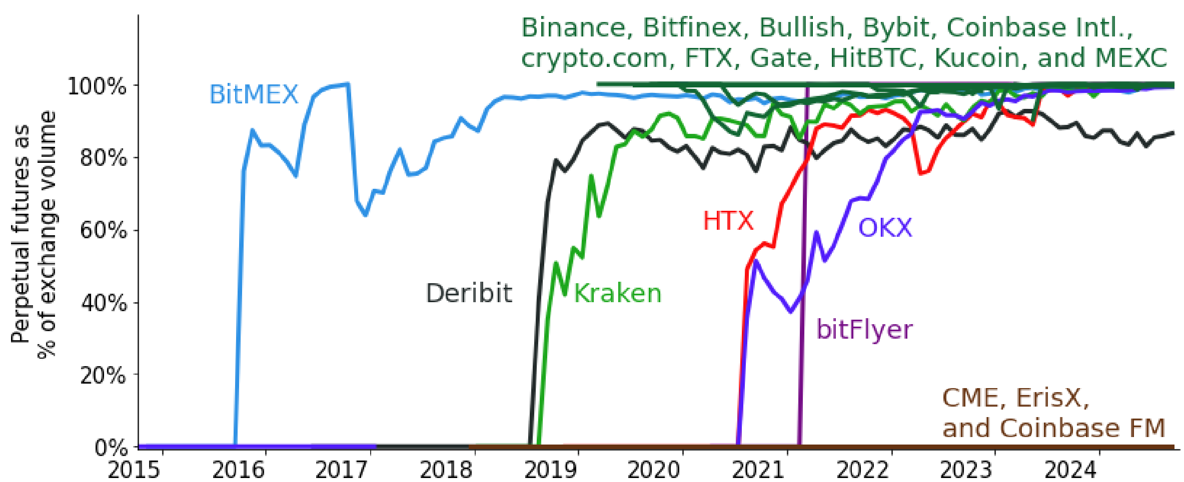
The early cryptocurrency futures were BTC-denominated bets on the USD/BTC price, as opposed to the more standard USD-denominated bets on the BTC/USD price.[6] This

[5]For example, the current Deribit BTC-USD perpetual future uses the mark price for F_t as discussed in footnote 2.3 sets λ so that the full difference between F and S is paid every eight hours, shrinks the result toward 0 by up to $0.003125\% S_t$, and bounds the result to $0.0625\% S_t$. Kraken applies similar caps, sets λ based on a one-day payout period, and sets the payment based on the average of $F_t^{REF} - S_t$ over the previous clock hour. BitMEX pays out every eight hours an amount based on the time-weighted difference over the prior eight hours, uses yet different bounds, shifts the rate up based on borrowing costs, and shrinks toward zero at a rate based on bid-ask spreads. He et al. (2022) discuss the mechanics of these contracts.

[6]Writing BTC-denominated contracts on USD was pinoered by ICBIT's founder Alexey Bragin to allow futures trading without reliance on fiat currency exchanges. Appendix Figure A5 provides the details of the first contract and Bragin (2015) provides more complete detail. To illustrate, suppose a trader goes long \$100 (4 BTC) of a futures contract at a price of \$25 per BTC and at settlement the BTC price is \$50. If she was long a conventional BTC/USD future as offered by the CME, she would have a profit of $\$100 = 4 \times (\$50 - \$25)$ because she had a USD-denominated long claim on 4 BTC each of which increased by \$25. If she was long a so-called inverse BTC future, she would have a profit of 2BTC$= \$100 \times (1/25 - 1/50)$ because she had a BTC-denominated short claim on \$100 and the dollar fell from 1/25th of a BTC to 1/50th.

structure allows futures trading to be purely BTC-based and independent from fiat currency. This market quirk is important to be aware of but has no impact on our results. We consider changes in USD-denominated portfolio values for all of our returns analysis and all of our figures are based on futures with the same form.

Figure 2: Rise of perpetual futures across exchanges.
This figure shows the percentage of reported futures volume accounted for by perpetual futures on the twenty cryptocurrency exchanges covered by Coin Metrics. These exchanges account for the vast majority of cryptocurrency futures volume. We use our scraped data for BitMEX, Kraken, and Deribit, with Coin Metric's data filling in for other exchanges.



3.2. Dominance of perpetual futures

Figure 2 shows how perpetual futures dominated cryptocurrency futures trading. It reports the portion of trading volume accounted for by perpetual futures for the twenty cryptocurrency futures exchanges covered by Coin Metrics, which together account for the vast majority of cryptocurrency futures trading. BitMEX was the first exchange to apply the perpetual futures contract to cryptocurrency. Major exchanges transitioned one by one in what appear to be regional waves: Europe-based Deribit and Kraken (then UK-based Crypto Facilities) saw their volume transition to perpetual futures in late 2018; China-based HTX and OKX saw a similar transition in mid-2020, potentially in response to competition from

China-based Binance; and Japan-based BitFlyer followed in early 2021.

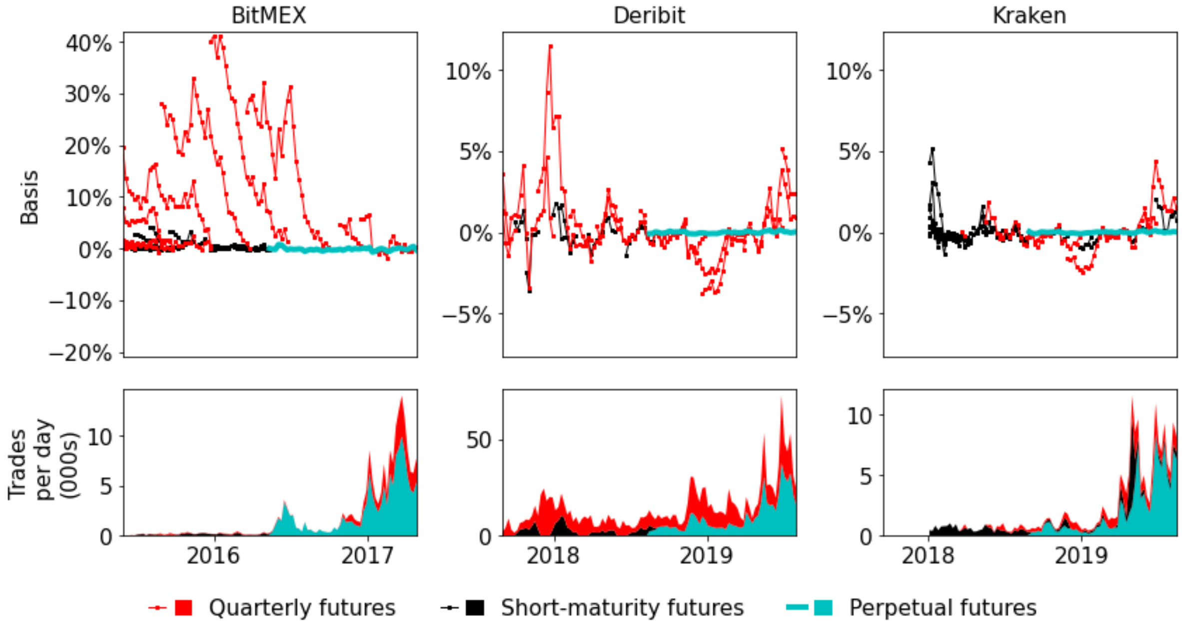
Perpetual futures' success is not confined solely to opaque offshore exchanges. They have eclipsed standard futures at the non-US futures arms of regulated exchanges (e.g., Kraken, Crypto.com, Coinbase International, Gemini) and the inherently transparent blockchain-based decentralized futures exchanges (e.g., Jupiter, dYdX, Hyperliquid). Further, Appendix Figure A2 shows that the same pattern emerges when we look at product offerings, which are not generally subject to manipulation.

The only exceptions to this pattern are the three exchanges offering cryptocurrency futures to Americans: the CME, Coinbase Financial Markets, and ErisX (owned by CBOE). This is not a coincidence. Although perpetual futures are not banned in the United States, their regulatory status is unclear. No perpetual futures products have been approved by the Commodity Futures Trading Commission (CFTC), the main US regulator and the current industry perception is that offering perpetual futures to Americans is not possible.

Figure 3 zeros in on BitMEX, Deribit, and Kraken which are the first earliest adopters and do not appear to manipulate data (see Section 2.1). For each exchange, we split futures into three classes: standard quarterly futures whose expirations fall on calendar quarters, short-maturity standard futures whose initial expirations are less than one quarter from the time of issue, and perpetual futures. The short-maturity standard futures had differing expirations on each exchange, with offerings including monthly, weekly, forty-eight-hour, and even daily maturities.

The top set of subplots shows the futures basis for each class of future. Keeping in mind that the scales used on different charts are different, several patterns are apparent. Quarterly futures prices fluctuate significantly from spot prices, while short-maturity futures prices are somewhat better behaved, and perpetual futures prices track the spot price extremely closely. In Bitmex's case, the basis of standard quarterly futures exceeds 40%, while the perpetual futures basis is consistently less than 1%. Deribit and Kraken transitioned during a relatively

Figure 3: Futures markets around perpetual futures introduction.
This chart documents features of the futures market on three cryptocurrency exchanges around the time of those exchanges' introduction of perpetual futures contracts. BTC futures are grouped into three classes: quarterly futures, perpetual futures, and short-maturity standard futures with monthly, weekly, forty-eight-hour, or daily maturities. The top subplots show the basis of each contract averaged to the week level, where the basis is the ratio of futures trade prices to the prices of their spot underlying. The bottom subplots report the daily fraction of trades accounted for by each type of contract.



tame period and see a quarterly futures basis that ranges from -4% to 8% and a perpetual futures basis that stays within a percentage point of zero.

The bottom set of subplots shows the number of trades of each contract.[7] Perpetual future's immediate dominance of trading volume is apparent. Further, for all three venues, the introduction of these futures is associated long-term volume increases, not just short-term cannibalization.

BTC futures have had a large positive basis at many points in their history, as shown by Appendix Figure A3. This positive basis means a simple 'cash and carry' trade of buying BTC and shorting its futures generates significant profit, as shown by academics (Christin

[7]We focus on the number of trades due to the extreme BTC price volatility during this period, the volume traded follows a similar pattern.

et al., 2022; Schmeling et al., 2023; He et al., 2022) and contemporaneous industry sources.[8] The consensus is that this anomaly was driven by speculative interest exceeding arbitrage capital. Other stories struggle to rationalize this extreme contango. Storage costs frequently drive the basis of commodities, however, storing BTC merely requires keeping a number (private key) or series of words (hash phrase) secret. Barriers to shorting the spot asset can rationalize a negative basis but not the pervasive positive basis we see. Credit risk is important in cryptocurrency, yet the right-way-risk of BTC futures predicts a high basis during panics and a low basis during good times, which is the precise opposite of what actually occurs (e.g., the FTX collapse in late 2022).[9]

Beyond creating an arbitrage opportunity, we argue that a high and volatile basis impedes trading and that eliminating this basis fueled the popularity of perpetual futures. Fundamentally, futures contracts that fluctuate dramatically from their underlying are worse products. Basis risk is one reason that market participants might dislike a large and variable futures basis. A speculator may want exposure to BTC's price movement and yet not have an opinion about the future speculative pressures and limits to arbitrage on a particular futures exchange, as might a hedger looking to offset BTC risk without realizing taxable capital gains. An arbitrager is able to hedge BTC risk (by buying BTC) but cannot hedge drawdowns created by extreme basis movements. Moreover, this arbitrager presumably prices in these crises risks; reducing the pricing efficiency of the standard future.

Pricing issues are another reason market participants might dislike a large and variable futures basis. A perpetual future's price is approximately reasonably well by the observable and tradable spot price, whereas the standard future's price also incorporates a significant

[8]For example, in 2015 the founder of BitMEX claimed that "One of the most common and profitable trading strategies when trading futures is cash and carry." in "Cash And Carry Arbitrage With BitMEX Futures," *BitMEX*.

[9]In particular, the risk of going short a BTC future is that the underlying exchange will not survive a drop in BTC. This type of wrong-way risk increases the futures price above the spot price. However, exchanges are most likely to fail during panics, not when times are good. The futures basis shows precisely the opposite pattern, being high in good times and negative when there are fears of exchange failure.

and relatively opaque basis component. A speculator who sees BTC trading for \$672 and expects its price to rise may (reasonably) be willing to buy a perpetual future that costs \$674 plus \$3 per day in financing costs until they unwind the position; but unwilling to buy a three-month standard future for \$877 (a 29% premium, as seen in the data) and unsure of the properties of that price. Similarly, a market maker trading BTC perpetual futures can focus on the BTC price, while making a market in longer maturity futures requires additionally predicting the imbalance between arbitrage capital and speculative demand.

We focus on channels associated with the basis of the future, but there are other potential explanations for perpetual future popularity, which we briefly comment on. First, it is possible that traders preferred contracts that did not need to be rolled over, which would reduce their holding costs. However, as previously discussed, these contracts emerged as part of a trend toward shorter and shorter maturities, not longer and longer maturities. Moreover, the holding period of crypto futures contracts on these exchanges is similar to that of the CME futures: short. For example, the turnover ratio for Sept. 2024 (daily volume to month-end open-interest) ranged from ranging for 0.31 (Deribit) to 1.49 (Kraken) for these exchanges, versus 1.26 for equity index futures, 0.39 for interest rate futures, and 0.17 for agriculture futures on the CME.[10]

Second, it is possible that perpetual futures contracts varied along aspects other than their perpetual structure. Beyond maturity, there are two salient aspects of futures exchanges: fees and margin requirements. For BitMEX and Kraken, fees were kept the same for the new perpetual futures contracts, and for Deribit, they were actually set higher than standard futures (perhaps explaining why perpetual futures are associated with increasing platform volume for Kraken and BitMEX but not Deribit). Margin requirements are the most salient contract term apart from fees: Kraken and Deribit launched perpetual futures with the same

[10]Crypto exchange data was retreived via CoinGecko. CME data was retreived via their disclosures for futures open interest and volume.

margin requirements as their standard futures but still saw perpetual futures capture most of the platform's volume shortly after launch. BitMEX offered lower margin requirements for short-maturity futures and perpetual futures than quarterly futures (25-50x); however, these higher margin requirements appear to have been driven in part by the poorly behaved prices of longer-maturity futures. A volatile futures basis impedes the liquidation of collateral.

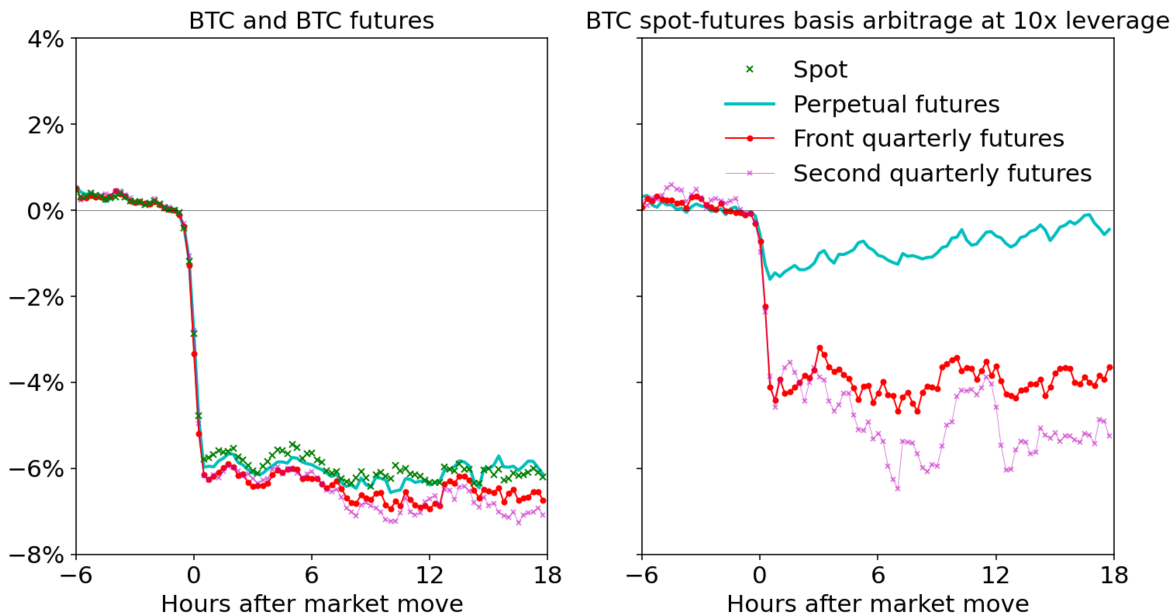
3.3. Perpetual futures are more stable in crises

We now turn to how different types of futures contracts perform when markets are under stress. Cryptocurrency markets are an ideal laboratory in which to study crises. The cryptocurrency market is frequently in situations that mirror financial crises, where there are fears of the solvency of major players, dramatic price movements, and scarce arbitrage capital. Further, most of the exchanges we looked at lacked any form of regulation for most of the sample, which not only fueled crises and balance-sheet uncertainties but meant that the market was not impacted by the prospect of moral-hazard-inducing bailouts.

The unregulated nature of cryptocurrency exchanges also gave them the freedom to write innovative contracts, unlike the tightly regulated traditional futures exchanges. This innovation was spurred by a high level of cross-exchange competition. In traditional markets, the trading of a certain type of future is generally concentrated on a single market. For example, the CME is home to precious metals trading, while the London Metals Exchange is home to industrial metals. Cryptocurrency futures exchanges typically offer similar contracts (futures on the price of BTC in USD) and are instead differentiated by linked product offerings and geographic focus. For example, Deribit focused on options, Kraken had USD markets and the approval of the FCA, bitFlyer dominated Japan, and BitMEX offered the best futures liquidity, but did not support fiat currency at all. This differentiation was driven both by regulation and by high switching costs.[11]

[11] To illustrate these costs, consider trading BTC futures on two exchanges. With a traditional broker, trading on the CME and CBOE is extremely straightforward. With Kraken and Binance, for example, one

Figure 4: Crisis returns of futures and hedged arbitrage strategies.
This figure shows the average cumulative returns of various futures strategies around the time of $> 5\%$ absolute spot price movements. We consider the BTC-USD set of BitMEX contracts for events in the 2015-2023 period where both perpetual futures and two quarterly futures were available. If multiple $> 5\%$ absolute changes occur, we consider the first such change in a 36-hour period. The left plot shows the cumulative returns of a strategy that holds \$100 spot BTC (using the average of Coinstamp and Bitfinex prices) and the cumulative returns to portfolios with the same BTC exposure constructed using BitMEX XBTUSD perpetual futures, front month (closest-maturing) futures, and second nearby (second closest-maturing) quarterly futures. The right panel shows the returns to arbitrage strategy that buys spot BTC and shorts the respective future with 10X leverage. Each line presents the cumulative cash-flows to the losing side of each trade, e.g., a long BTC trade around price decreases and a short BTC trade around price increases. We plot 15-minute averages of 1-minute cross-scenario medians.



We first look at the performance of quarterly and perpetual futures during crypto's many crises. Figure 4 assesses the risk created by basis movements, focusing on the cash flows created by margin, futures, and arbitrage trades. We consider BTC contracts on BitMEX because it offers the longest time series and considers periods where we have perpetual futures

needs separate margin balances in separate currencies custodied at the two exchanges (Binance is primarily USDT based, Kraken is USD based), slow onchain transfers between exchanges, potentially the loss of steep volume discounts (Kraken fees start at 2bp (5bp) for makers (takers) and fall to −0.6bp (1bp) for traders with high previous-month volume as shown at Kraken Futures Fees), and the purchase large quantities of exchanges tokens (Binance offers large fee discounts to holders of its BNB token).

and two quarterly futures. We look at the hourly absolute spot price changes of more than 5%. These movements would have imposed significant stress on many market participants, as BTC accounted for most of the market cap of cryptocurrencies throughout this sample. We consider the losing side of each trade, for example, for price increases (decreases), we consider the returns to a short (long) BTC positions. This reflects the fact that many cryptocurrency trading institutions have short futures positions hedged with either spot assets or CME futures.

The left plot considers the returns of holding a constant value of BTC. We show the returns from holding spot, the cash flows from holding BitMEX's original XBTUSD perpetual future, and the cash flows from both the front (closest maturity) or second nearby (second closest maturity) of BitMEX's quarterly futures.[12] For futures contracts, these returns correspond to margin calls. By construction, each of the four strategies suffers significant losses. The spot strategy's immediate peak losses in the two hours around the price jump are the smallest, averaging 5.8 percentage points. For reference, the worst days of the 2008 financial crisis saw daily price declines of 7-8 percentage points.

The futures strategies suffer additional losses when the futures basis moves against them: the perpetual future losses are 3% (0.1 percentage points) larger than the spot losses, the front-quarterly losses are 8% (0.5 percentage points) larger, and the second nearby quarterly's losses are 10% (0.6 percentage points) larger. Given the quarterly futures have larger absolute basis, it is intuitive that they also lose more when the market turns.

A limits to arbitrage story is the most natural explanation for these patterns. Price decreases are associated with negative speculators sentiment, the liquidation of the most optimistic speculators, and a deleveraging of market makers. These changes cause the futures basis to move with the price of the underlying spot asset. The similar losses on the two

[12]This chart is created using 10% annual interest on the margin position used to purchase the spot BTC, however, these costs are negligible at the daily level.

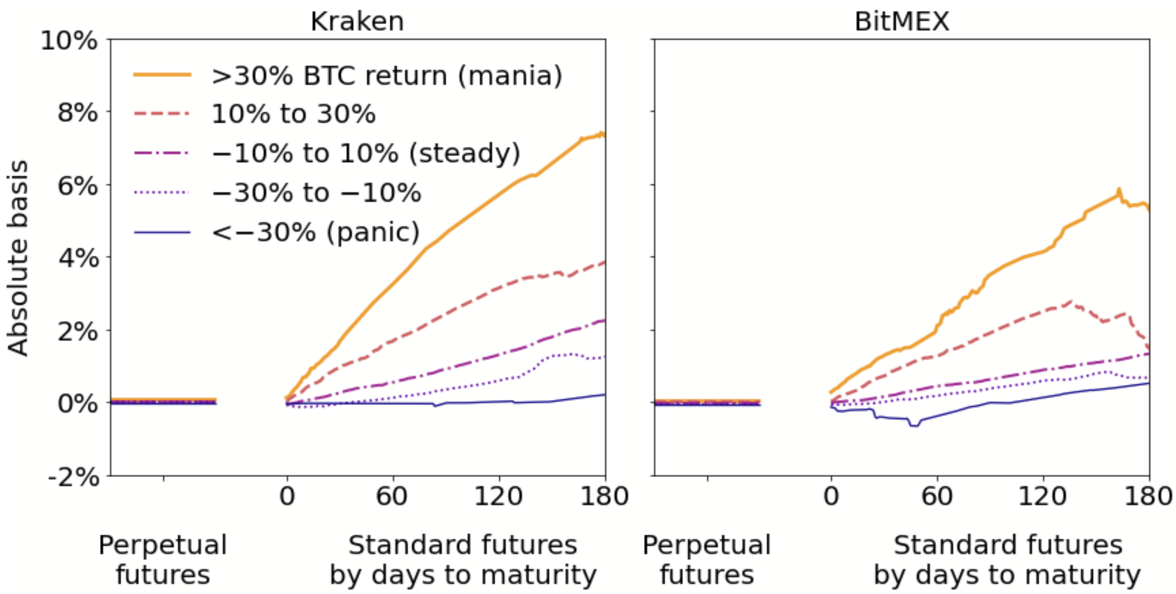
quarterly futures are consistent with an expectation that basis risk will subside in the medium term. Further, the similarity in those strategies' losses is inconsistent with a simple liquidity premium story as the second quarterly future is dramatically less liquid than the front quarterly future, yet their basis change is similar.

The right plot shows the losses of BTC basis arbitrage trades where futures positions are hedged using offsetting spot positions. These arbitrage strategies follow a similar structure to treasury cash-futures basis trades or covered interest parity trades. To allow us to use the same axis, and to reflect the prevalence of leverage in basis trades (e.g., Siriwardane et al. (2022)), we have scaled these arbitrage trades by 10 to reflect hypothetical 10x leverage. In the long term, these arbitrage strategies should be near riskless as their eventual payoff is fixed. In practice, we see marked deviations in the same way the literature has shown marked covered interest parity deviations. Quarterly futures show larger losses, with a levered-basis arbitrage using the second nearby quarter losing up to 4.6% in the two hours around the shock, the front quarter strategy losing 4.3%. The perpetual future sees a far lower drawdown, with peak losses of just 1.5%. Not only are the peak losses from the perpetual futures arbitrage less than half as severe, the losses are also shorter lived: twenty-four hours after the shock, the perpetual futures arbitrage has erased all of the losses, while the quarterly futures strategies are close to their crisis lows.

Looking at the futures curve in different market states offers direct evidence of the link between basis and risk. Figure 5 shows the basis of perpetual and standard futures, with standard futures being plotted with respect to their time to maturity. We use the recent change in the spot BTC price as a proxy for market sentiment. We divide our sample into five regimes based on where the spot BTC price sits relative to its 30-day exponential moving average: below -30% (panic), -30% to -10%, -10% to 10% (steady), 10% to 30%, and $> 30\%$ (mania).

Three patterns are apparent. First, the basis of perpetual futures is close to zero in each

Figure 5: BTC futures basis by tenor and regime.
This chart shows the median basis of BTC futures under different regimes on the Kraken (left) and BitMEX (right) across days in our trade-level data where a future with more than 120 days to expiration was offered. Regimes are defined based on the previous day's BTC price relative to its 30-day exponential weighted moving average. For each day and each exchange, we calculate the interpolated basis at each daily maturity using linear interpolation and extrapolate up to seven days past the shortest or longest maturity. Each point on the chart is the median of the interpolated basis values for days in a given regime.



regime while the basis of delivery futures is large.

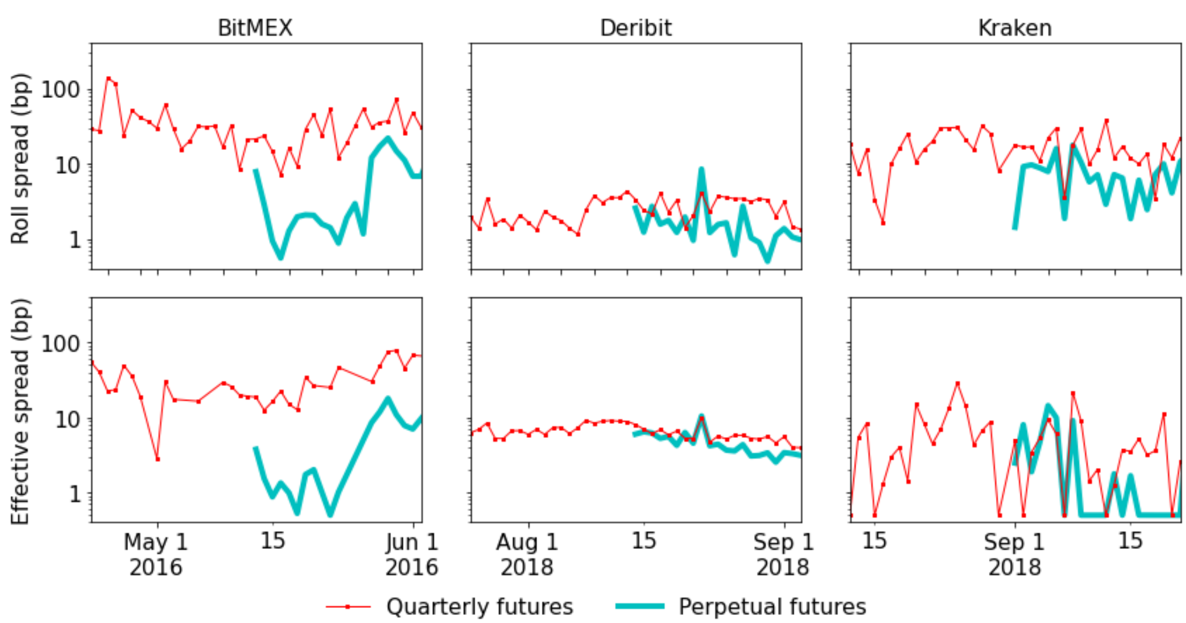
Second, the magnitude of the basis of standard futures increases with tenor or time to maturity. The futures basis is near zero for standard futures close to maturity, while it ranges from -1% to 4% for standard futures 90 days to maturity on both exchanges and 0% to $6-7\%$ for standard futures 180 days to maturity, consistent with Rutledge (1976) and Castelino and Francis (1982).

Third, the basis of standard futures is high after past BTC spot price increases (e.g., on average 4% for 90 days-to-expiration futures on Kraken during manias) and low after BTC price decreases (e.g., -1% during panics). This pattern means that the basis changes of delivery futures tend to magnify the gains or losses from the underlying price movements, consistent with Figure 4. For example, if the price of BTC increases, a trader with a short position in the future will lose both from the futures price increase and because the basis increases.

3.4. Perpetual futures have better market quality

It is natural to think that reducing basis risk would improve price efficiency. Fluctuations in the futures basis not only create additional risk for market makers, they introduce additional noise into pricing. Throughout, we will try to control for the fact that perpetual futures are currently far more popular than delivery futures.

Figure 6: Futures market quality following perpetual future introduction.
This figure covers days surrounding the introduction of perpetual futures for BitMEX, Deribit, and Kraken and shows the liquidity of newly introduced perpetual future and the front quarterly future. The Roll spread is twice the square root of the autocorrelation of relative price changes. The effective spread is the average basis points liquidity takers overpay relative to fair prices. For BitMEX, we set the fair price equal to the midprice immediately before the trade. For Deribit and Kraken, we set the fair price to be the exchange-provided mark-to-market price used for liquidations, which are exponential moving averages of midpoint prices, adjusted for movements in the underlying asset. For this chart only, we set the effective spread equal to 1/2 of a basis point for days where the effective spread is negative.



We first present suggestive evidence from when these contracts were first introduced.

Figure 6 presents the Roll spreads and effective spreads on perpetual and quarterly futures for our three early adopters, with lower values denoting better liquidity. The day-to-day volatility in these measures is primarily measurement error, in particular for Kraken which had fewer daily trades.

Across exchanges and measures, we see perpetual futures offering better liquidity. The differences between perpetual and standard futures are statistically significant for each exchange and each measure. Their economic magnitudes vary, with BitMEX seeing dramatic improvements and Deribit and Kraken seeing more muted improvements. The degree of basis risk present in standard futures is a potential explanation for this differential effect. As Figure 3 shows, BitMEX introduced perpetual futures during a period when the futures basis was high and volatile and saw substantially all the volume move to these contracts, with liquidity on standard futures drying up. Kraken and Deribit show more muted improvements, potentially because these exchanges introduced perpetual futures during a time when the futures basis was lower and less volatile.

Table 1 expands our sample to include six major cryptocurrencies and provides data on these liquidity measures for different futures. Given the extreme skewness of the underlying distributions, we Winzorize at the 1% level for all measures. The first columns summarize our measures for the sample. Perpetual futures have higher volume and better liquidity across venues however these differences are not statistically significant due to the high skewness of the underlying distribution.

The next set of columns uses matching to control for the fact perpetual futures have far more volume. For each day and each standard future on that day, we find a lower volume perpetual future traded on the same coin and the same venue with the closest number of trades (dropping perpetual futures with $>$ 20% more trades or $<$ 30% fewer trades. In practice, that means matching a front-quarter standard future with a perpetual future future on a less-liquid underlying. For example, matching a front quarter BTC-USD future

Table 1: Market quality for matched perpetual and quarterly futures.
This table reports market quality metrics for standard quarterly futures and perpetual futures on BitMEX, Deribit, and Kraken. A data point is a symbol-day, considering only six liquid cryptocurrencies (BTC, ETH, SOL, XRP, BCH, LTC). The first columns report the entire sample for these coins for quarterly and perpetual futures. Diff. presents the difference between the two types of futures. ***, **, and * denote that this difference is statistically significant at the 1%, 5%, and 10% levels, respectively, for standard errors based on coin-month clustering. The next set of columns repeats this exercise for a sample matched on volume, coin, and month. Each quarterly future symbol-day is paired with the perpetual future which has a number of trades closest to the quarterly future's number of trades, considering only matches within venue-coin-month and provided there exists a matching future with 70-120% the number of trades. We drop trading days with fewer than 10 trades and winzorize at the 1% level for all measures.

	All symbol-days			Volume matched symbol-days		
	Quarterly	Perpetual	Diff.	Quarterly	Perpetual	Diff.
BitMEX						
Observations	14055	14241	186	125	125	0
Distinct coins	5	6	1	5	5	0
Distinct symbols	156	16	-140	51	11	-40
Daily trades	3039.05	12948.0	9908.94	6568.97	6635.5	66.53
Effective spread (bp)	8.42	2.75	-5.67	8.83	2.63	-6.21***
Roll spread (bp)	10.17	4.4	-5.76	11.34	4.62	-6.72***
Kraken						
Observations	9798	12963	3165	235	235	0
Distinct coins	5	6	1	5	5	0
Distinct symbols	126	12	-114	100	10	-90
Daily trades	723.38	3448.41	2725.03	1879.21	1864.36	-14.85
Effective spread (bp)	11.74	6.56	-5.18	16.12	6.65	-9.46***
Roll spread (bp)	13.62	8.61	-5.02	20.88	8.51	-12.38***
Deribit						
Observations	11318	6707	-4611	117	117	0
Distinct coins	3	6	3	3	3	0
Distinct symbols	53	9	-44	38	6	-32
Daily trades	3274.68	13112.26	9837.58	14273.81	14086.04	-187.77
Effective spread (bp)	7.47	6.14	-1.33	8.55	4.01	-4.55***
Roll spread (bp)	5.17	3.55	-1.62	6.81	2.45	-4.36***

with a BTC-JPY perpetual. In this sample, the perpetual futures have similar volumes by construction, yet we see perpetual futures demonstrating far better liquidity. There are statistically significant and economically meaningful differences across measures: across the three venues, the effective spread and roll spread are 53-71% lower.

Table 2: Market quality for matched perpetual and delivery futures.
This table presents determinants of the spread for a sample of futures contracts. The dependent variable is effective spread after adding 1 and log transforming for Specifications (1) to (5) and is a similar measure for the Roll spread for Specification (6). *Standard future* denotes futures with expiration dates (as opposed to perpetual futures with no expirations and funding payments). *Maturity* is the log of one plus the contract's maturity in days, where the maturity of perpetual futures is assumed to be 0. *Trades* is the log of the number of trades on a given day. *Inverse*, *BTC-quoted*, and *Quanto* are contract features. Exchange corresponds to the exchange data source. Standard errors are clustered at the coin-month level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Effective spread					Roll Spread
	(1)	(2)	(3)	(4)	(5)	(6)
Standard future	0.86***	0.77***	0.65***	0.88***		0.66***
	(0.06)	(0.21)	(0.05)	(0.14)		(0.06)
Maturity					0.18***	
					(0.02)	
Trades		-0.05	-0.04**	-0.00	-0.01	-0.07***
		(0.05)	(0.02)	(0.04)	(0.03)	(0.02)
Inverse		-3.14***	0.14**	-0.20*	-0.22	0.01
		(0.41)	(0.06)	(0.12)	(0.19)	(0.11)
BTC-quoted		-0.32		-0.89***	-0.28	-0.20
		(0.38)		(0.29)	(0.26)	(0.12)
Quanto		-0.29			-0.39	0.03
		(0.53)			(0.42)	(0.22)
Exchange	All	BitMEX	Deribit	Kraken	All	All
Coin-month-exchange FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,817	16,506	8,870	13,441	38,817	38,817
R^2	10.16%	8.55%	46.93%	9.33%	10.62%	19.13%

Table 2 uses a regression specification to test for liquidity differences. We use a panel OLS regression on the log of the effective spread in basis points plus 1. We use the logged value both to allow us to talk about relative increases and to prevent seldom-traded coins from driving our results. Each data point is from a symbol s on coin c traded on day d in month m on exchange e, according to the following specification:

$$\log\left[EffectiveSpread_{s,d} + 0.01\%\right] = X_{s,d}^{\top}\beta + \eta_{e,c,m} + \epsilon_{s,d}. \tag{6}$$

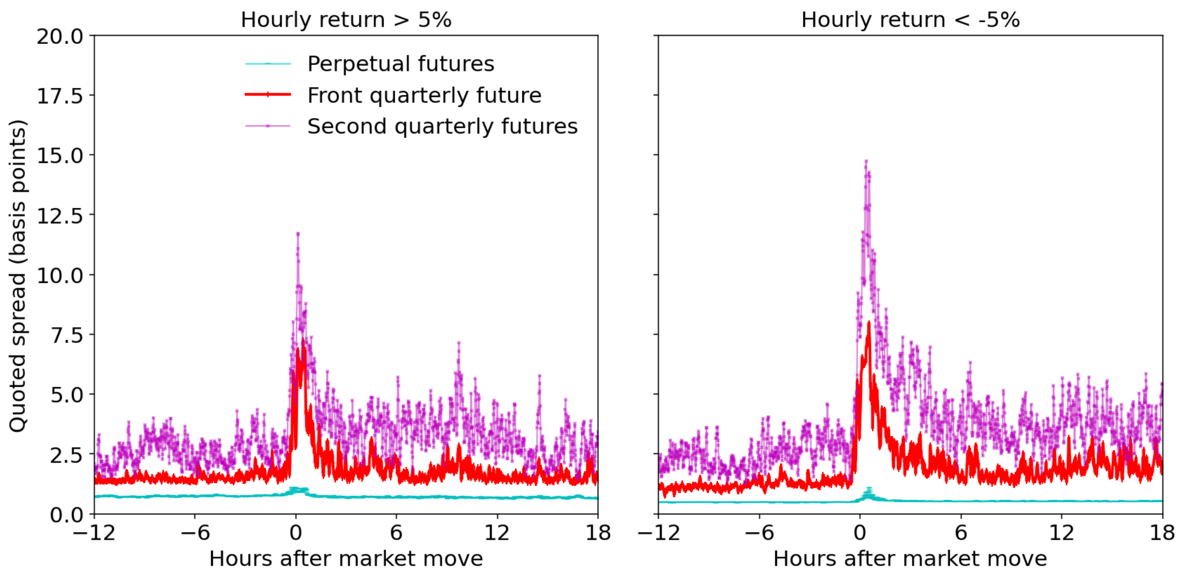
Our independent variables X_{s_d} are characteristics of symbol s on day d. The variable of interest is either an indicator for the symbol being a standard future (as opposed to a perpetual future) or the log of one plus the number of days until the future's expiration, with perpetual futures set to 0. Control variables are the log of the number of trades on that symbol-day and indicator variables for contract features including the symbol being an inverse, quanto, or BTC-quoted future. All specifications include exchange-coin-month $\eta_{e,c,m}$. Standard errors are presented based on clustering at the coin-month level.

Specification (1) confirms that standard futures indeed have significantly (0.91 log points) higher spreads. Specifications (2) through (4) add controls for trades and contract characteristics and repeat that test on each of our three venues, yielding similar results. Specification (5) considers time to maturity instead of an indicator for being a standard future, and shows a similar pattern of longer maturity contracts being less liquid. Specification (6) shows that similar results hold when using the Roll liquidity measure as the dependent variable.

Given our focus on limits to arbitrage, it is natural to ask how different styles of futures contract perform during market turbulence. However, as Table 1 and Table 2 show, the normal-times liquidity of these contracts is very different which makes statistical comparisons challenging. Figure 7 shows the quoted (bid-ask) spread of BitMEX BTC-USD futures around large hourly moves, in a similar manner to Figure 4. We consider the periods around $>$ 5% hourly increases (left) and $<$ 5% hourly decreases (right) and report the median figure median bid-ask spread across all such periods, averaged to the half-hour level.

All contracts see a large increase in quoted spread around the time of these extreme price moves. Consistent with challenges in pricing and managing basis risk around these

Figure 7: Quoted spreads of futures around the time of price jumps.
This figure shows the median quoted spread of various types of futures around the time of $> 5\%$ hourly spot price increases (left plot) and $< -5\%$ spot price decreases (right panel). We consider the BTC-USD set of BitMEX contracts for events in the 2015-2023 period where both perpetual futures and two quarterly futures were available. If multiple $> 5\%$ changes occur, we consider the first such change in a 36-hour period. Each line presents the quoted (bid-ask) spread for a type of contract and a certain time relative to the large hourly move. We plot rolling 5-minute averages of 1-minute cross-scenario medians.



jumps, perpetual futures see much smaller spread increases. Specifically, perpetual futures quoted spreads increase to an average of 0.66 bp in the two hours surrounding positive price jumps, 34% higher than the average 0.49 bp peak in the 1-12 hours prior to the jumps. The perpetual future show similar patterns for negative price jumps, with spreads increasing 25% from 0.49 bp to 0.66 bp. The front quarterly future sees much larger increases in spreads in both relative and absolute terms for both the price increases (peak quoted spread increases 161% from 1.66 bp to 4.36 bp) and price decreases (increases 249% from 1.44 bp to 5.03).

The second quarterly future has spreads that are about twice as large as the front quarter future prior to the shock (3.19-3.39 bp) and sees these spreads increase less than the front quarter future in relative terms (115-161% versus 161-249%) and more in absolute terms (142-210 versus 61-112 bp). Given the second quarterly future has much worse ex-ante liquidity, it is perhaps surprising that the increases around shocks are at all comparable. This suggests that the better crisis performance of the perpetual future may not be solely driven by its better overall liquidity: the perpetual future has far better performance than the somewhat-less-liquid front quarterly future, while the front quarterly future performs has only slightly better relative performance than the somewhat-less-liquid second quarterly future.

Unfortunately, the data do not allow us to test whether this basis risk is priced. First, the fact that the vast majority of volume uses perpetual futures has presumably made noise trader basis risk less severe. Second, we have only a short time series of data over which BTC has appreciated dramatically. Betting against the basis is akin to betting against future intense BTC speculation (because the basis covaries with returns), which was ex-post a losing proposition as speculative interest in BTC remained at a very high level.

4. A Theoretical Model of Noise Trader Risk

Consider an asset with fundamental value and an associated futures contract on that asset. Assume that the asset price S_t follows a diffusion process:

$$dS_t = rdt + \sigma_S dY_t^S, \tag{7}$$

where $r \geq 0$ is the constant risk-free rate, and Y_t^S is a Brownian motion.

Let P_t be the price of the futures contract and $B = P_t - S_t$ be its basis. We consider a general contractual form that allows us to model both standard and perpetual future contracts. Let T be the expiration of the futures contract, at which time the futures contract settles for S_T. Let δ be the funding rate of that futures, so that there is a continuous payment of δB from the long side of the contract to the short side. Under this setup, a standard future has a finite expiration date $T < \infty$ and no funding rate $\delta = 0$, while a perpetual future contract has a positive funding rate $\delta \geq 0$ and an infinite horizon $T = \infty$.

The futures contract is of 0 total supply and traded by n^S mass of liquidity traders S, a unit mass of long-term hedgers H, and n^M mass of market makers M. Each agent's wealth, W_i, evolves based on their holdings of the futures contract x_t^i:

$$dW_t^i = x_t^i dP_t - x_t^i \delta dt. \tag{8}$$

Liquidity traders get an additional benefit from holding the asset of $x^i G_t^S dt$. We call $G_t^S dt$ the liquidity trading factor and it drives trade. This value represents a persistent factor affecting liquidity traders' valuation of the asset, such as a liquidity constraint, hedging need, or sentiment.[13]

We model the liquidity trading factor as following an Ornstein-Uhlenbeck (OU) process, a continuous time analogy of the $AR(1)$ process:

$$dG_t^S = -\psi G_t^S dt + \sigma_G dY_t^G, \tag{9}$$

where Y_t^G is an independent Brownian motion. The parameter ψ characterizes the mean reversion speed of the liquidity trader's preference, and σ_G represents the rate of its diffusion.

Subject to this preference, the liquidity traders and market makers optimize their have mean-variance preferences over their instantaneous wealth (and benefit) changes:

$$dW_t^i - rW_t^i + G_t^S x^i \mathbb{I}[i = S]dt - \frac{\gamma^i}{2} Var(dW_t^i). \tag{10}$$

The variance penalty can be interpreted as risk aversion or inventory costs. We use γ^M to denote the risk aversion parameter for market makers, and γ^S for liquidity traders. Long-

[13] An equivalent interpretation is that liquidity traders have different beliefs from the market maker regarding S_t, with the liquidity traders believing the underlying asset price follows $dS_t = (G_t^S + r)dt + \sigma_S dY_t^S$. Under this interpretation, G_t^S represents the disagreement between market makers and liquidity traders.

term hedgers do not optimize and have constant aggregate hedging demand x^H.

We assume that the future's basis B depends on time t and liquidity trading factor G_t^S. Following the standard assumption in continuous time analysis, we further assume that $B = B(G_t^S, t)$ is differentiable in t and twice differentiable in G_t^S.

Now we are ready to define equilibria.

Definition 4.1 (Equilibrium)**.** An equilibrium is a characterization of future basis function $B(G_t^S, t)$, the liquidity trader's strategies x_t^S, and the market markers strategies x_t^M such that:

1. Given $B(G_t^S, t)$, market makers pick their futures contract position x_t^M to maximize their utilities;
2. Given $B(G_t^S, t)$, liquidity traders pick their futures contract position x_t^S to maximize their utilities;
3. Market clearing: $x^H + n^M x_t^M + n^S x_t^S = 0$ for $\forall t \in [0, T]$.

We next present our main result. All proofs can be found in Appendix B.

Proposition 4.2. *There exists an equilibrium such that $B = k(t)G_t^S + j(t)x^H$. In this equilibrium, the absolute basis $|B|$ is decreasing in the funding rate δ and the liquidity trading factor mean reversion parameter ψ.*

Proposition 4.2 shows that there exists an equilibrium such that the basis $B = k(t)G_t^S + j(t)x^H$ is a linear function of the liquidity trading factor G_t^S and the long-term hedging demand x^H. Specifically, we refer to $k(t)G_t^S$ as the liquidity part of the basis and $j(t)x^H$ as the hedging part. In the proof, we derive a closed-form expression for the liquidity component:

$$k(t) = \underbrace{\frac{1/\gamma^S}{1/\gamma^S + 1/\gamma^M}}_{\text{Proportional risk-aversion}} \underbrace{\frac{1}{\psi + \delta}\left(1 - e^{(\psi+\delta)(t-T)}\right)}_{\text{Liquidity factor risk}}. \tag{11}$$

Similarly, we can solve $j(t)$ given the the boundary condition $j(T) = 0$ and $(j' - \delta j)(n^S\gamma^M + n^M\gamma^S) + \gamma^M\gamma^S(\sigma_G^2 k^2(t) + \sigma_S^2) = 0$:

$$j(t) = \underbrace{\frac{1}{1/\gamma^S + 1/\gamma^M}}_{\text{Aggregate risk-aversion}} \underbrace{\frac{\sigma_G^2 k^2(t) + \sigma_S^2}{\delta}\left(1 - e^{\delta(t-T)}\right)}_{\text{Hedger price impact scalar}}. \tag{12}$$

When $\delta = 0$, $j(t)$ converges to $\frac{\sigma_G^2 k^2(t)+\sigma_S^2}{n^S/\gamma^S + n^M/\gamma^M}(t-T)$. For standard delivery future contracts,

$$P_t = S_t + \frac{1/\gamma^S}{1/\gamma^S + 1/\gamma^M}\frac{1}{\psi}\left(1 - e^{\psi(t-T)}\right)G_t^S + \frac{\sigma_G^2 k^2(t) + \sigma_S^2}{n^S/\gamma^S + n^M/\gamma^M}(t-T)x^H. \tag{13}$$

In the case of perpetual future contract,

$$\begin{aligned} P_t =& S_t + \frac{1/\gamma^S}{1/\gamma^S + 1/\gamma^M}\frac{1}{\psi + \delta}G_t^S \\ &+ \frac{1}{n^S/\gamma^S + n^M/\gamma^M}\frac{\sigma_G^2 k^2(t) + \sigma_S^2}{\delta}x^H. \end{aligned} \tag{14}$$

The absolute liquidity basis part $k(t)$ is increasing (decreasing) in γ^M (γ^S). The degree of risk aversion affects each group's trading behavior, hence their market power. When market makers are risk neutral, they will trade until the basis $B(G_t^S, t) = 0$. When liquidity traders are less risk-averse, they trade more, making the liquidity factor more influential in future prices. The absolute basis is decreasing in both δ and ψ because a more persistent liquidity factor (lower ψ) means a larger price impact from the current liquidity shock of liquidity traders, but a high funding rate normalizes that.

We can further obtain equilibrium positions x_t^i:

$$x_t^S = -\frac{1}{n^S\gamma^M + n^M\gamma^S}(\gamma^M x^H - \frac{n^M G_t^S}{\sigma_G^2 k^2(t) + \sigma_S^2}), \tag{15}$$

and

$$x_t^M = -\frac{1}{n^S\gamma^M + n^M\gamma^S}(\gamma^S x^H + \frac{n^S G_t^S}{\sigma_G^2 k^2(t) + \sigma_S^2}). \tag{16}$$

The next result states which future contracts are more favored by traders.

Corollary 4.3. *In equilibrium, the instantaneous trading volume is increasing in both the funding rate δ and liquidity trading factor mean reversion parameter ψ. The subjective utilities for liquidity traders and market makers are increasing in funding rate δ. In particular, given a delivery contract with an expiration horizon T, there exists $\overline{\delta}$ such that both liquidity traders and market makers prefer the perpetual future contract if and only if $\delta \geq \overline{\delta}$.*

Because the hedging demand x^H is a constant, trades happen between market makers and liquidity traders, driven by the evolution of liquidity trading factor G_t^S. A smaller absolute basis implies a low future price volatility, making agents more willing to trade. As a result, the instantaneous trading volume $|\frac{dn^S x_t^S}{dt}|$ is increasing in both δ and ψ. Adding a funding rate (δ) mitigates the impact of liquidity trading factor shocks, restoring trading towards the no-noise trader risk optimum. In equilibrium, both market makers and liquidity traders prefer high funding rates and are willing to trade more.

The last result of the proposition elaborates on two channels that determine the basis. For standard delivery future contracts, the basis converges to 0 as the time t approaches the expiration date T. For the perpetual future contract, there is no expiration date so the basis never converges to 0. However, the funding rate δ forces the future price to approach the fundamental asset value, shrinking the basis. Our analysis shows that when the funding rate is high enough, it can reduce the basis to be lower than the average basis for standard delivery contracts.

We then examine the hedger's preference when they roll over their position using future contracts of horizon T. To characterize hedgers' utility, we further assume that the hedge position x^H is made up of x^H mass of hedgers who each hold unit exposure, are born at a uniform rate x^H/l, live l, and have mean-variance preferences over the tracking error within their lifetime. These hedgers roll over fixed maturity futures contracts at their expiration as needed to keep their exposure constant. Then the following corollary confirms that hedgers prefer perpetual future contracts as well.

Corollary 4.4. *Hedger welfare is higher with an appropriately set perpetual future than any fixed maturity standard future.*

Intuitively, a longer rollover horizon increases the expected average basis, making hedging more costly. On the other hand, a higher funding rate lowers the basis, making hedging less costly. The corollary above suggests that when the funding rate is sufficiently high, hedgers prefer perpetual future contracts over standard delivery contracts.

Combining corollary 4.3 and 4.4, we can conclude that the introduction of perpetual future contracts increases both trading volume and social welfare for all three types of traders in the market.

Our last result concerns the change of the speculator population. Given the high volatility in the Crypto market, liquidity traders are often wiped out of the market, resulting in a large drop of n after a dramatic price move. The following corollary studies the effect of massive speculative position liquidation on future market trading.

Corollary 4.5. *The absolute speculative basis decreases in* $\frac{n^M}{n^S}$.

When the market power of liquidity traders drops, the speculative demand in the future market becomes less volatile and the absolute basis shrinks. Consequently, market makers can afford a larger inventory capacity and are willing to trade with a bigger open interest in the future market.

5. Conclusion

Perpetual futures follow their underlying due to frequent "funding rate" payments between the long and short side of the contract. These funding rate payments mean the payoff to a perpetual future is similar to very short-maturity delivery futures that are repeatedly rolled over into other short-maturity futures. This structure transfers basis risk (the risk of the futures price moving away from its spot underlying) from arbitrageurs to other market participants, in the same manner a LIBOR-linked loan transfers interest rate risk away from banks.

We use data from several cryptocurrency venues to show that perpetual futures outcompeted delivery futures, reduce basis risk, and have better crisis market quality. The magnitudes are large: perpetual futures account for almost all the volume of ex-US future exchanges, reduce the losses of futures basis arbitrage trades by half, and see lower transaction cost increase during market stress.

Our model rationalizes these facts and shows that perpetual futures are better able to suit the needs of a variety of market participants. Capital-constrained intermediaries prefer them because reducing basis risk reduces their crisis losses. Reducing the costs of intermediaries increases arbitrage pressure and reduces the price hedgers pay. Hedgers also benefit from reduced risk overall, even if they have long-term cash flows. This risk reduction occurs because, for the average hedger, the risk of higher-than-expected funding payments is less than the basis risk incurred when rolling over hedges or exiting hedges at dates other than standard futures maturities.

The CFTC has not approved perpetual futures, leading to a general perception that these contracts are effectively banned in the US. Although the cryptocurrency market differs from traditional financial markets in many ways, perpetual futures success and better crisis performance suggest banning them from traditional markets is unwarranted.

References

Acharya, V. V., Lochstoer, L. A., and Ramadorai, T. (2013). Limits to arbitrage and hedging: Evidence from commodity markets. *Journal of Financial Economics*, 109:441–465.

Akey, P., Robertson, A., and Simutin, M. (2022). Noisy factors. *Rotman School of Management Working Paper Forthcoming.*

Alexander, C., Deng, J., and Zou, B. (2021). Hedging with bitcoin futures: The effect of liquidation loss aversion and aggressive trading. *arXiv preprint arXiv:2101.01261.*

Avalos, F. and Sushko, V. (2023). Margin leverage and vulnerabilities in us treasury futures. *BIS Quarterly Review.*

Bai, J. and Collin-Dufresne, P. (2019). The cds-bond basis. *Financial Management*, 48(2):417–439.

Barth, D., Kahn, R. J., et al. (2021). Hedge funds and the treasury cash-futures disconnect. *OFR WP*, pages 21–01.

Binsbergen, J. v., Brandt, M., and Koijen, R. (2012). On the timing and pricing of dividends. *American Economic Review*, 102(4):1596–1618.

Bragin, A. (2015). Inverse futures in bitcoin economy. *Available at SSRN 2713755.*

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies.*

Castelino, M. G. and Francis, J. C. (1982). Basis speculation in commodity futures: The maturity effect. *The Journal of Futures Markets (pre-1986)*, 2(2):195.

Chen, J., Lin, D., and Wu, J. (2022). Do cryptocurrency exchanges fake trading volumes? an empirical analysis of wash trading based on data mining. *Physica A: Statistical Mechanics and its Applications*, 586:126405.

Choi, K. J., Lehar, A., and Stauffer, R. (2022). Bitcoin microstructure and the kimchi premium. *Available at SSRN 3189051*.

Christin, N., Routledge, B. R., Soska, K., and Zetlin-Jones, A. (2022). The crypto carry trade. *CMU Working Paper*.

Cong, L. W., Li, X., Tang, K., and Yang, Y. (2022). Crypto wash trading. Technical report, National Bureau of Economic Research.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.

Deng, J., Pan, H., Zhang, S., and Zou, B. (2020). Minimum-variance hedging of bitcoin inverse futures. *Applied Economics*, 52(58):6320–6337.

Du, W. and Tepper (2018). Deviations from covered interest rate parity. *The Journal of Finance*, 73(3):915–957.

Duffie, D. and Rahi, R. (1995). Financial market innovation and security design: An introduction. *Journal of Economic Theory*, 65(1):1–42.

Figlewski, S. (1984). Hedging performance and basis risk in stock index futures. *The Journal of Finance*, 39:657–669.

Foucault, T., Sraer, D., and Thesmar, D. J. (2011). Individual investors and volatility. *The Journal of Finance*, 66(4):1369–1406.

Gehr Jr, A. K. (1988). Undated futures markets. *Journal of Futures Markets*, 8(1):89–97.

Glicoes, J., Iorio, B., Monin, P., and Petrasek, L. (2024). Quantifying treasury cash-futures basis trades. *FEDS Notes*, 3(2024):8–3.

Hazelkorn, T. M., Moskowitz, T. J., and Vasudevan, K. (2020). Beyond basis basics: Liquidity demand and deviations from the law of one price. Technical report, National Bureau of Economic Research.

He, S., Manela, A., Ross, O., and von Wachter, V. (2022). Fundamentals of perpetual futures. *arXiv preprint arXiv:2212.06888*.

Miller, S. M. (1986). Financial innovation, depository-institution deregulation, and the demand for money. *Journal of Macroeconomics*, 8(3):279–296.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance*, 39(4):1127–1139.

Rutledge, D. J. (1976). A note on the variability of futures prices. *The Review of Economics and Statistics*, pages 118–120.

Schmeling, M., Schrimpf, A., and Todorov, K. (2023). Crypto carry. *Available at SSRN 4268371*.

Shen, J., Yan, H., and Zhang, J. (2014). Collateral-motivated financial innovation. *The Review of Financial Studies*, 27(10):2961–2997.

Shiller, R. J. (1993). Measuring asset values for cash settlement in derivative markets: hedonic repeated measures indices and perpetual futures. *The Journal of Finance*, 48(3):911–931.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance*.

Simsek, A. (2013). Financial innovation and portfolio risks. *American economic review*, 103(3):398–401.

Siriwardane, E., Sunderam, A., and Wallen, J. L. (2022). Segmented arbitrage. Technical report, National Bureau of Economic Research.

Tufano, P. (2003). Financial innovation. *Handbook of the Economics of Finance*, 1:307–335.

Appendix A Additional tables and figures

Figure A1: Futures trading on the same underlying with varying expirations. This figure shows the popularity of different futures tenors for a typical non-roll period day. Futures are split into the most popular maturity (e.g., the nearest expiring), the second most popular (e.g., the second nearest expiring), and the third most popular (e.g., the third nearest expiring), and all other contracts. Data from the CME for the Euro/USD contract (E6), the 10-year U.S. treasury note (ZN), the S&P 500 E-mini (ES), Henry Hub natural gas (NG), West Texas Intermediate crude oil (CL), gold (GC), high-grade copper (HG), corn (ZC), and live cattle (LE). Contracts on metrics such as the VIX or SOFR are not shown because different maturities correspond to fundamentally different underlyings.

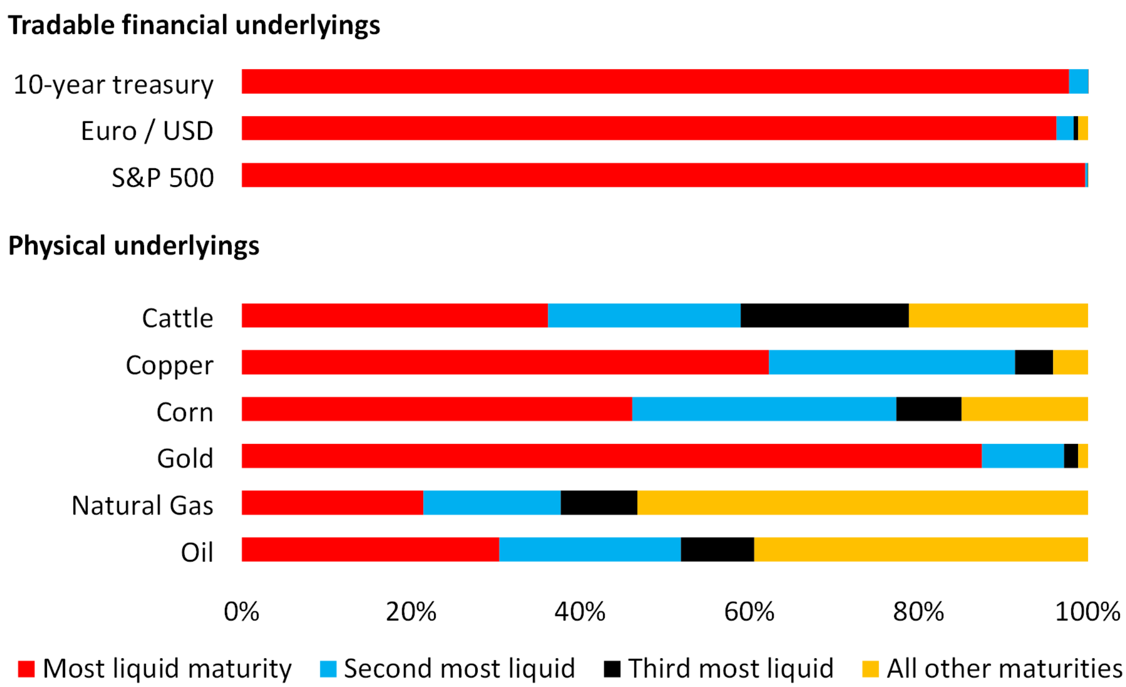


Figure A2: Rise of perpetual future product offerings across exchanges.
This figure shows the percentage of futures product symbols accounted for by perpetual futures on the twenty cryptocurrency exchanges covered by Coin Metrics. Each maturity of a standard future accounts for a separate symbol. These exchanges account for the vast majority of cryptocurrency futures volume. We use our scraped data for BitMEX and Kraken Futures and Coin Metric's data for other exchanges.

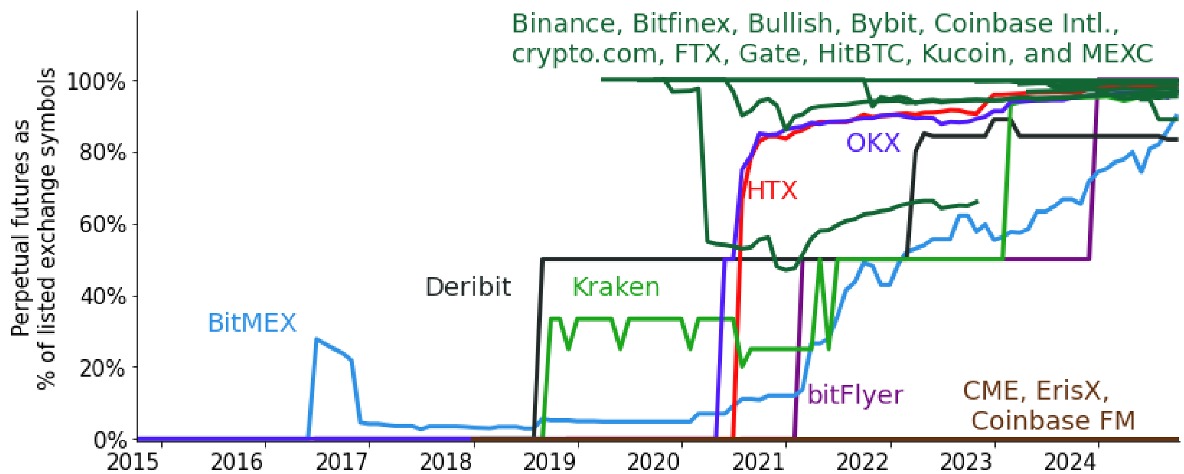


Figure A3: Cryptocurrency futures prices relative to spot prices.
This figure shows the daily average futures basis for the nearest two expiring quarterly BTC-USD futures and the perpetual BTC-USD future on the cryptocurrency exchange BitMEX. The basis for each trade is calculated as the ratio of the futures price to the price of the underlying spot asset, minus one. The spot price is the simple average of the last traded price on Bitfinex, Coinbase, and Coinstamp. The futures price is the mid-point at the end of the minute. The plotted basis is the average of the minutes in two-week periods.

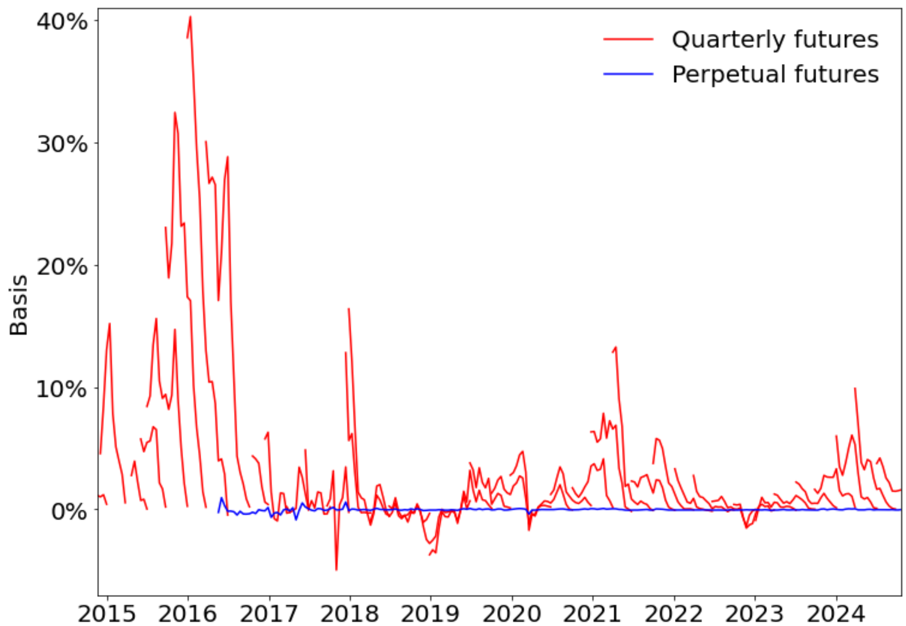


Figure A4: Contractual evolution of BTC futures.
This chart summarizes the futures products offerings of exchanges offering crypto futures prior to Dec. 2018. Product offering data are sourced from the Wayback Machine - Internet Archive, following a procedure akin to Akey et al. (2022) and by scraping the longest-living BTC forum, BitcoinTalk.org. This chart considers only futures contracts and does not consider exchanges offering margin loans for spot trading (e.g., the early days of Bitfinex) or bilateral bets (e.g., over-the-counter contracts for difference or TeraExchange). This chart excludes ErisX which offered futures trading starting 2019 and was acquired by the CBOE in 2021.

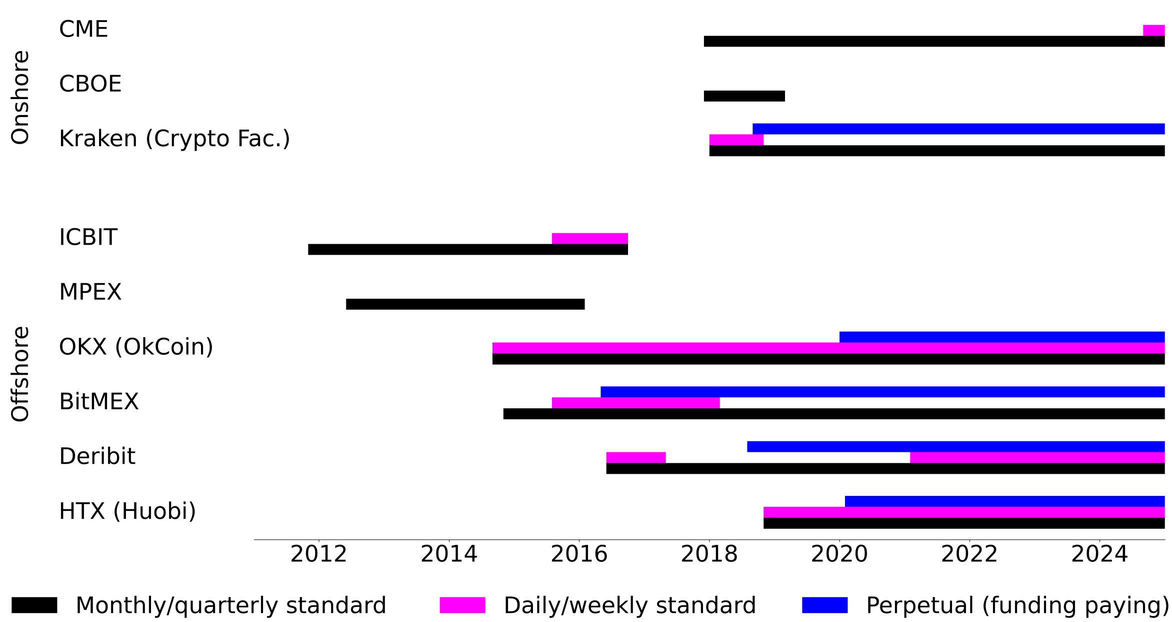


Figure A5: Contract specifications for the first BTC inverse futures contract.

BTCUSD-12.12

Full symbol: BTCUSD-12.12
Short symbol: BUZ2
Name: Futures contract on the Bitcoin - US dollar exchange rate.
Type: Settled in BitCoins.
Lot size: 10 USD.
Quotes: In US dollars.
Minimal price step (W): 0.0001 USD
Cost of the minimal price step (R): -0.001 USD
Trading starts: 06.2012
Last trading day: 15.12.2012
Settlement date: 15.12.2012
Settlement: Positions are settled based on the volume weighted average rate of USD/BTC on the exchange with the most average monthly volume (for the month of contract settlement) during the contract settlement day by transferring variation margin between contract holders.
Maximum price change within one trading session: 10% in each direction relative to the close price of the last trading session.

Settlement price: $13.4870

Variation margin is calculated as follows:
*VM = (1/PriceClose - 1/PriceOpen) * W/R;*

Example:
Joe opened his position by buying 5 contracts at 5.3172. He was lucky and in a few hours price went up to 5.8769 where he sold all his 5 contracts. His profit is:
*Profit = ((1 / 5.8769 - 1 / 5.3172) * -10) * 5 = 0.89556 BTC*
Using anology to spot trading, that would mean that Joe wanted to speculate on the BTC/USD rate, and had spare $50. He bought Bitcoins spending $50 when the price was $5.3172 for 1 BTC, and then sold as much Bitcoins as it was needed to get his $50 back. However, since Bitcoin rose in value, his accounts now have 0.89556 BTC as profit plus $50 he had before.

Unlike spot trading, in order to do the same operation in the futures market, Joe needs to have about 3.4 BTC in his account as initial margin.

Source: Wayback Machine - Internet Archive

Appendix B Proofs

B.1. Proof of Proposition 4.2

Proof. For an agent of type $i \in \{M, S\}$ and her time t position of future contract x_t^i, her wealth evolves as:

$$\begin{aligned} dW_t^i =& x_t^i dS_t + x_t^i dB\left(G_t^S, t\right) - x_t^i \delta B\left(G_t^S, t\right) dt \\ =& x_t^i[(G_t^i + r)dt + \sigma_S dY_t^S + B_t(.)dt - \psi G_t^S B_G(.)dt + \frac{1}{2}\sigma_G^2 B_{GG}(.)dt \\ & + \sigma_G B_G(.)dY_t^G - \delta B(.)dt], \end{aligned} \tag{17}$$

where $B_t(.)$ is the first order derivative with respect to t, $B_G(.)$ is the first order derivative with respect to G_t^S, and $B_{GG}(.)$ is the second order derivative with respect to G_t^S.

For an agent of type i, she chooses the optimal x_t^i to maximize her expected preference:

$$E[dW_t^i - rW_t^i - \frac{\gamma^i}{2} Var(dW_t^i)]. \tag{18}$$

Substituting equation (17), one can solve the optimal x_t^i, for $i \in \{M, S\}$:

$$x_t^i = \frac{G_t^i + B_t(.) - \psi G_t^s B_G(.) + \frac{1}{2}\sigma_G^2 B_{GG}(.) - \delta B(.)}{\gamma^i\left(\sigma_G^2 B_G^2(.) + \sigma_S^2\right)}. \tag{19}$$

The market clearing condition $x^H + n^S x_t^S + n^M x_t^M = 0$ suggests that

$$x^H\left(\sigma_G^2 B_G^2(.) + \sigma_S^2\right) + (\frac{n^M}{\gamma^M} + \frac{n^S}{\gamma^S})[B_t(.) - \psi G_t^s B_G(.) + \frac{1}{2}\sigma_G^2 B_{GG}(.) - \delta B(.)] + \frac{n^S}{\gamma^S} G_t^S = 0. \tag{20}$$

We conjecture $B\left(G_t^S, t\right) = k(t)G_t^S + j(t)x^H$. Then we have $B_t(.) = k'(t)G_t^S + j'(t)x^H$; $B_G(.) =$

$k(t)$; and $B_{GG}(.) = 0$. Substituting those and canceling out terms, we have

$$\begin{aligned} 0 =& [n^S\gamma^M + (n^S\gamma^M + n^M\gamma^S)(k' - \psi k - \delta k)]G^S \\ &+ [(j' - \delta j)(n^S\gamma^M + n^M\gamma^S) + \gamma^M\gamma^S(\sigma_G^2 k^2 + \sigma_S^2)]x^H. \end{aligned} \tag{21}$$

Because in equilibrium the market clear condition always holds for any G_t^S and x^H. We can solve $k(t)$ and $j(t)$ separately. Start with $k(t)$, given the boundary condition $k(T) = 0$ (because basis converges to 0 as t approaches T), and $n^S\gamma^M + (n^S\gamma^M + n^M\gamma^S)(k' - \psi k - \delta k) = 0$, one can solve the ODE:

$$k(t) = \frac{n^S\gamma^M}{n^S\gamma^M + n^M\gamma^S}\frac{1}{\psi + \delta}\left(1 - e^{(\psi+\delta)(t-T)}\right). \tag{22}$$

Similarly, we can solve $j(t)$ given the the boundary condition $j(T) = 0$ and $(j' - \delta j)(n^S\gamma^M + n^M\gamma^S) + \gamma^M\gamma^S(\sigma_G^2 k^2 + \sigma_S^2) = 0$:

$$j(t) = \frac{\gamma^M\gamma^S(\sigma_G^2 k^2 + \sigma_S^2)}{(n^S\gamma^M + n^M\gamma^S)\delta}(1 - e^{\delta(t-T)}). \tag{23}$$

Thus,

$$P_t = S_t + \frac{n^S\gamma^M}{n^S\gamma^M + n^M\gamma^S}\frac{1}{\psi + \delta}\left(1 - e^{(\psi+\delta)(t-T)}\right)G_t^S + \frac{\gamma^M\gamma^S(\sigma_G^2 k^2 + \sigma_S^2)}{(n^S\gamma^M + n^M\gamma^S)\delta}\left(1 - e^{\delta(t-T)}\right)x^H. \tag{24}$$

To show that the absolute basis is decreasing in both ψ and δ, we only need to show that both $k(t)$ and $j(t)$ are (weakly) decreasing in both ψ and δ. That is equivalent to show that $\frac{1}{x}(1 - e^{-x(T-t)})$ is decreasing in x, for $\forall x > 0$. Taking the first order derivative with respect to x, we obtain:

$$\frac{d}{dx}\frac{1}{x}(1 - e^{-x(T-t)}) = \frac{1}{x^2}[(1 + x(T-t))e^{-x(T-t)} - 1]. \tag{25}$$

We then want to show that $(1+x(T-t))e^{-x(T-t)}-1<0$. To achieve this, notice that $(1+x(T-t))e^{-x(T-t)}-1=0$ when $x=0$, and taking the first order derivative with respect to x again, we have $(T-t)e^{-x(T-t)}-(T-t)[1+x(T-t)]e^{-x(T-t)}<0$, for $\forall x>0$. $\square$

B.2. Proof of Corollary 4.3

Proof. Given the equilibrium price equation 24 and market clear condition, we can further obtain x_t^i:

$$x_t^S=-\frac{1}{n^S\gamma^M+n^M\gamma^S}(\gamma^M x^H-\frac{n^M G_t^S}{\sigma_G^2k^2+\sigma_S^2}), \tag{26}$$

and

$$x_t^M=-\frac{1}{n^S\gamma^M+n^M\gamma^S}(\gamma^S x^H+\frac{n^s G_t^S}{\sigma_G^2k^2+\sigma_S^2}). \tag{27}$$

We then show that the trading volume is increasing in ψ and δ. Because x^H is a constant, the instantaneous trading volume can be characterized as:

$$\begin{aligned}|\frac{dn^Sx_t^S}{dt}|=&\frac{n^Sn^M}{(n^S\gamma^M+n^M\gamma^S)(\sigma_G^2k^2+\sigma_S^2)}|\frac{dG_t^S}{dt}|\\=&\frac{n^Sn^M}{(n^S\gamma^M+n^M\gamma^S)(\sigma_G^2k^2+\sigma_S^2)}|-\psi G_t^Sdt+\sigma_GdY_t^G|.\end{aligned} \tag{28}$$

Since $k(t)$ is decreasing in both ψ and δ, the instantaneous trading volume increasing in both ψ and δ.

From the proof of Proposition 4.2, the subjective expected utility of speculators S is

$$\begin{aligned}E[\frac{\gamma^S\,(\sigma_G^2B_G^2(.)+\sigma_S^2)}{2}(x_t^S)^2]=&QE[(\gamma^Mx^H-\frac{n^MG_t^S}{\sigma_G^2k^2+\sigma_S^2})^2]\\=&Q[(\gamma^Mx^H)^2+(\frac{n^M}{\sigma_G^2k^2+\sigma_S^2})^2\frac{\sigma_G^2}{2\psi}],\end{aligned} \tag{29}$$

where $Q\equiv\frac{\gamma^S\left(\sigma_G^2B_G^2(.)+\sigma_S^2\right)}{2(n^S\gamma^M+n^M\gamma^S)}$ is a constant. Because $k(t)>0$ is decreasing in δ, specualtor S's subjective expected utility is increasing in δ. Similarly, one can prove the same result for market maker M.

To prove the last result, first notice that we only need to show that for any standard delivery contract, there exists $\overline{\delta}$ such that the all perpetual future contracts with a refunding rate higher than $\overline{\delta}$ will have a lower $k(t)$. It is straightforward to see that when $\delta = \overline{\delta} \equiv (\frac{1}{1-e^{\psi(t-T)}} - 1)\psi$, the perpetual future contract and the standard delivery future share the same basis. Then the perpetual future contracts have a lower absolute basis and higher open interest if and only if $\delta \geq \overline{\delta}$. $\square$

B.3. Proof of Corollary 4.4

Proof. Let τ be the random birth time. We reuse $k(t)$ and $j(t)$ to denote $k(t \bmod T)$ and $j(t \bmod T)$. Let H^τ denote G shifted by the hedger's random birth time, $H^\tau_s = G_{s+\tau}$. An agent born at time τ almost surely has total cash flow

$$\pi(\tau, \{H^\tau\}) = H^\tau_l k(\tau + l) + j(\tau + l)h - H^\tau_0 k(\tau) - j(\tau)h$$

$$- \sum_{n \in \mathcal{Z}: 0 < Tn - \tau < l} \left(H^\tau_{Tn-\tau} k(0) + j(0)h\right) - \int_\tau^{\tau+l} \delta \left(k(\tau + t)H^\tau_t + j(\tau + t)h\right) dt$$

Use $^-$ to denote the realized means of variables between 0 and T. Then

$$E[\pi(\tau, \{H^\tau\})|\{H^\tau\}] = \frac{1}{T}\int_0^T \pi(\tau, \{H^\tau\})dt = \left(H^\tau_l - H^\tau_0\right)\bar{k} - l\left(\frac{1}{T}\left(\bar{H}k(0) + j(0)h\right) + \delta\left(\bar{k}\bar{H} + \bar{j}h\right)\right)$$

Which allows us to calculate unconditional moments

$$E[\pi] = E[E[\pi_\tau|\{H\}]] = -lh\left(\frac{1}{T}j(0) + \delta\bar{j}\right)$$

$$Var[\pi] = Var[E[\pi_\tau|\{H\}]] + E[Var[\pi_\tau|\{H\}]]$$

$$\geq Var[E[\pi_\tau|\{H\}]] = Var\left[\left(H^\tau_l - H^\tau_0\right)\bar{k} + \bar{H}\frac{l}{T}\left(k(0) + \delta\bar{k}\right)\right]$$

with that inequality binding for perpetual futures.

Fix T for the delivery future and consider a perpetual future with δ' such that

$$\frac{1}{\psi}\left(1-\frac{1-e^{-\psi T}}{\psi T}\right)=\frac{1}{\psi+\delta'},$$

letting $'$ denote this future's properties. By construction, we have $k(0)>\bar{k}\geq\bar{k}'$ and so $\frac{1}{T}j(0)>\delta j'$. Thus, this perpetual futures contract has a higher expected cash flow without higher variance. $\square$

B.4. Proof of Corollary 4.5

Proof. From the proof of corollary 4.2, $k(t)$ can be rewritten as

$$\begin{aligned} k(t)=&\frac{n^S\gamma^M}{n^S\gamma^M+n^M\gamma^S}\frac{1}{\psi+\delta}\left(1-e^{(\psi+\delta)(t-T)}\right)\\ =&\frac{1}{1+\frac{n^M}{n^S}\frac{\gamma^S}{\gamma^M}}\frac{1}{\psi+\delta}\left(1-e^{(\psi+\delta)(t-T)}\right). \end{aligned} \tag{30}$$

Then $k(t)$ is decreasing in $\frac{n^M}{n^S}$. $\square$