Competition in the Cryptocurrency Exchange Market*

Junyi Hu[†], Anthony Lee Zhang[‡]

June 2025

Abstract

Cryptocurrency exchange market structure is fragmented, since cryptocurrencies are fungible, but customers cannot move freely across cryptocurrency exchanges. We build a model where exchanges with captive customers are linked by arbitrageurs, showing that "star-shaped" equilibria can exist, in which arbitrageurs endogenously coordinate on one exchange as a liquidity hub. The model predicts that large exchanges' listing decisions should influence price dispersion, arbitrage flows, trade volumes, and listing decisions on small exchanges. We provide evidence for the model's predictions using data on exchange prices and trade volumes, as well as blockchain token flows across exchanges.

^{*}We thank Ben Golub, Ralph Koijen, Andreas Park, Daniel Rabetti, Tim Roughgarden, Julia Selgrad, Taisiya Sikorskaya, Amir Sufi, Simon Trimborn, Yizhou Xiao, Xian Wu, Yuanyuan Wang, the wagmi chat, and seminar participants at CICF 2024, CUHK, EFA 2023, the Hong Kong Conference for Fintech, AI, and Big Data, UTS, Loyola University, USI Lugano, the University of Zurich, the University of Geneva, Luohan Academy, and the Stevanovich Center at UChicago for helpful comments. We are especially grateful to Jinfei Sheng for comments that greatly improved the paper. We are grateful to Hrishikesh Relekar, Yijing Ren, Haichuan Wang, and Chupei Zhang for excellent research assistance.

[†]Columbia University, Columbia Business School; jh4736@gsb.columbia.edu

[‡]University of Chicago, Booth School of Business; anthony.zhang@chicagobooth.edu.

1 Introduction

Markets typically function best when they are centralized, yet modern financial asset markets are often surprisingly fragmented. Fragmentation is particularly extreme in the international market for cryptocurrency exchanges. Cryptocurrencies are very concentrated as an asset class: a few cryptocurrencies account for over half of the entire sector's market cap. But cryptocurrency exchanges are very fragmented: there are over a thousand active cryptocurrency exchanges worldwide, offering their customers fairly similar sets of assets to trade.

This paper analyzes the fragmented structure of the international cryptocurrency exchange market. We ask a number of questions. Why are there so many cryptocurrency exchanges? How do exchanges strategically interact with each other? More tentatively, is the current market structure of crypto exchanges a historical accident, or is it in some sense adapted to primitive features of cryptocurrencies as an asset class?

To motivate our answer, we first observe that crypto exchange markets resemble the structure of *commodity* markets. Commodities such as gold, oil, or wheat are globally fungible, but costly to transport. They are thus not traded on centralized global exchanges. Instead, buyers and sellers within regions are matched in small regional markets; supply and demand imbalances across regions are addressed by specialized "arbitrageurs", who transport goods across markets in response to price gaps. Certain regional markets may evolve into trading "hubs", serving as focal points of arbitrage activity and distribution centers for surrounding markets.

We propose that similar forces shape cryptocurrency exchange market structure. Cryptocurrencies are globally fungible, but crypto customers in different regions may use different fiat currencies and face different regulatory restrictions; this effectively creates trading costs across customer groups, analogous to physical transportation costs for commodities. A natural response to these costs is a "network of exchanges" market structure, in which many small exchanges emerge to connect customers who are "close" to each other, and specialized arbitrageurs take on the role of moving inventory between exchanges.

Building on these ideas, in this paper, we analyze a simple model in which a number of ex-ante identical exchanges with captive customer bases are imperfectly linked by arbitrageurs. We show that a *star network* can emerge in equilibrium, where arbitrageurs each link a "peripheral" exchange to a common "central" exchange. In such equilibria, the central exchange endogeneously acquires outsized influence: since peripheral exchanges are connected through the center, listings of new coins on central exchanges will tend to decrease price dispersion and increase arbitrage flows and trading volumes across all peripheral exchanges. We find

empirical support for our model's predictions. We thus provide a narrative explaining how crypto exchange fragmentation emerges from the primitives of cryptocurrencies as an asset class, which has implications for the regulation of exchanges and inter-exchange arbitrageurs.

In our model, a single coin is traded on a number of ex-ante identical exchanges, each with a fully captive customer base. Customers have declining marginal utility for the coin due to inventory costs, and receive random inventory shocks. A number of identical arbitrageurs each choose a pair of exchanges to link. Arbitrageurs cannot hold net positions in the coin, but can transport quantity between exchanges subject to transportation costs, profiting from price gaps between linked exchanges. Markets clear in a one-shot uniform-price double auction, in which we assume customers and arbitrageurs both bid competitively. In equilibrium, arbitrageurs move the coin from exchanges with larger inventory shocks towards those with smaller shocks. Arbitrage is imperfect due to transportation costs, so exchanges with excess initial inventory end up with lower final prices.

Our core result is that, when arbitrageurs' transportation costs are low relative to captive customers' inventory costs, star network equilibria emerge, in which arbitrageurs coordinate on each linking a "peripheral" exchange to a common "central" exchange. Intuitively, the central exchange is deeper but has more stable prices, creating a tradeoff between exploiting large initial price gaps and reducing price impact. When transportation costs are high, arbitrageurs are capacity-constrained and prioritize larger initial price gaps, favoring peripheral-peripheral connections. When transportation costs are low, arbitrageurs prefer the deeper liquidity available at the central exchange, allowing star network equilibria to emerge.

Our model makes a number of empirical predictions. Prices and arbitrage flows should reflect the star structure of arbitrage connections: peripheral exchange prices should tend to be more correlated with the central exchange than with each other, and arbitrage flows should tend to concentrate at the central exchange. Since peripheral exchanges are connected through their links to the central exchange, central exchanges' decisions to list new coins should have a disproportionate effect on market outcomes. Before a coin can be traded on a central exchange, each peripheral exchange is an isolated market for the coin: price dispersion across peripheral exchanges should be high, and trade volume on each exchange should be low. After central exchanges list a coin, arbitrageurs can move inventory across peripheral exchanges in response to peripheral inventory shocks; this should decrease coin price dispersion, and increase trade and arbitrage volumes on peripheral exchanges. Anticipating increased trade volumes, peripheral exchanges which have not already listed the coin should have a higher propensity to list.

We test our model's predictions using price and volume data for 770 coins across 254

exchanges from January 2017 to December 2023, as well as a novel dataset of cross-exchange coin flows on the Ethereum blockchain. Following the literature and industry discourse, we treat two exchanges, Binance and Coinbase, as central. As our model predicts, centrality is visible in both prices and arbitrage flows. Price correlations show a star network structure: price correlations between peripheral exchanges and Coinbase or Binance are systematically higher than price correlations between pairs of peripheral exchanges. Moreover, consistent with findings in the literature (Kroeger and Sarkar, 2017; Makarov and Schoar, 2020; Dyhrberg, 2020; Borri and Shakhnov, 2023), we find that price gaps between peripheral and central exchanges tend to quickly mean-revert, with peripheral prices moving more than central prices. As the literature has documented using Bitcoin blockchain data (Makarov and Schoar, 2021), Ethereum on-chain flows also have a core-periphery structure: arbitrage flows through Binance and Coinbase are large in absolute terms, as well as relative to their trade volume.

We then test our model's predictions about the effects of central exchange listings by running difference-in-differences regressions. After central exchanges list a new coin, dispersion in coin prices across peripheral exchanges tends to decrease; trade and arbitrage volumes involving the coin tend to increase; and peripheral exchanges that have not listed the coin have a higher propensity to list. These results are consistent with our model's predictions. They contrast with a view in which customers can easily move across exchanges; in this case, central exchange coin listings should tend to cannibalize peripheral exchanges.

Consistent with the view that these effects are driven by peripheral-central arbitrage links, we find that peripheral exchanges which have greater price correlations with central exchanges also experience larger volume increases after central exchange listings, and are more likely to follow central exchange listings. Our results hold throughout our sample period, and survive in subsamples of exchanges that implement AML/KYC requirements, and that are less likely to fake trade volumes. We provide evidence suggesting our results are not fully explained by confounding factors such as changes in investor attention to coins, or possible "endorsement" effects of central exchange listings.

Our theoretical and empirical results suggest that large exchanges have substantial influence over the rest of the cryptocurrency market; policymakers may wish to monitor or regulate the listing decisions of large exchanges. Our results suggest that *arbitrageurs* are also a point of potential regulatory oversight. Policymakers could impose registration and reporting requirements on crypto exchange arbitrageurs; similar to the requirements imposed on market participants playing analogous roles in traditional financial markets, such as authorized participants in ETF markets, and designated market makers in equity markets. Arbitrageur regulation is a potential policy lever regulators can use to modulate the efficiency of local cryptocurrency markets.

Finally, our results provide a narrative explanation of modern crypto exchange market structure. Cryptocurrency exchanges are the boundary between the world's unified international cryptocurrency rails, and its many fragmented fiat and regulatory systems. As in commodity markets, a "network of local exchanges" loosely linked by arbitrage is a natural response. Extrapolating this narrative, we can tentatively predict the near-term future of exchange market structure: some degree of exchange fragmentation is likely to persist as long as fiat and regulatory frictions across customer bases remain.

This paper theoretically and empirically analyzes the structure of the cryptocurrency exchange market. Theoretically, we show that factors analyzed in prior empirical studies – customers who wish to trade identical assets, separated by frictions that resemble "transportation costs" – can support a core-periphery network structure in equilibrium, through the strategic behavior of arbitrageurs. In such a structure, "central" exchanges' listing decisions should influence cross-exchange price dispersion, peripheral exchanges' trade and arbitrage volumes, and listing decisions. We empirically confirm these predictions; to our knowledge, our findings that central exchange listings decrease peripheral price dispersion, and increase arbitrage volumes and listing propensities, are new to the literature.

Our theoretical results are related to classic papers showing how liquidity-seeking traders may concentrate on a single market (Admati and Pfleiderer, 1988; Pagano, 1989), as well as more recent results on core-periphery network formation in OTC markets (Chang and Zhang, 2015; Wang, 2016; Babus and Hu, 2017; Hugonnier, Lester and Weill, 2022; Farboodi, Jarosch and Shimer, 2023) and interbank lending markets (Craig and Ma, 2022; Farboodi, 2023), and the effects of core-periphery market structure on market outcomes (Viswanathan and Wang, 2004; Dunne, Hau and Moore, 2015; Üslü, 2019; Dugast, Üslü and Weill, 2022). We believe we are the first paper to show conditions under which core-periphery networks of *exchanges* can emerge, through arbitrageurs' strategic link formation decisions.

Our model contrasts with a theoretical literature on exchange competition, which shows that competition tends to lower fees, but may lower execution likelihood (Colliard and Foucault, 2012) or increase adverse selection or "quote sniping" opportunities (Budish, Cramton and Shim, 2015; Biais, Foucault and Moinas, 2015; Baldauf and Mollner, 2021). The economics of our model is different because we assume exogeneously captive customers. Technically, instead of the limit-order-book models used in this literature (Biais, Foucault and Moinas, 2015; Chao, Yao and Ye, 2019; Budish, Lee and Shim, 2024), our model is a simplified demand-schedule submission game (Rostek and Yoon, 2020).

Malamud and Rostek (2017), Rostek and Yoon (2021), Wittwer (2021), and Chen and Duffie (2021) use double-auction models to study how fragmentation affects trading outcomes and social welfare. In contrast, our goal is to analyze the endogeneous formation of interexchange links. Babus and Kondor (2018) model dealer-customer OTC trading on an exogeneous network as a one-shot demand-schedule game; we differ from their model in many ways, such as assuming private values and studying arbitrageurs' entry decisions. Other papers on endogeneous market structure in double-auction models are Yoon (2018), Babus and Parlatore (2022), and Wu (2024).

Empirically, this paper is related to a large literature showing that there is substantial price dispersion across cryptocurrency exchanges, which appears to be driven by demand shocks that are not fully eliminated by arbitrage. Makarov and Schoar (2020) show that inter-exchange BTC price differences are too large to be explained purely by transaction costs, and argue they are instead driven by capital controls. Financial restrictions appear to be an important driver of crypto demand: Yu and Zhang (2022) show that local BTC prices increase with economic policy uncertainty, and Choi, Lehar and Stauffer (2022) find that crypto prices tend to be higher in more financially restrictive countries.

Dyhrberg (2020), Borri and Shakhnov (2023), Hautsch, Scheuch and Voigt (2024), and other papers document a variety of frictions to inter-exchange arbitrage: price dispersion is related to exchanges' geographic location, fiat currency choices, BTC price volatility, and settlement latency. Augustin, Rubtsov and Shin (2023) show that the introduction of BTC futures contracts decreases exchange price dispersion. Other papers analyzing crypto price dispersion across exchanges include Kroeger and Sarkar (2017), Makarov and Schoar (2019), Tsang and Yang (2020), Borri and Shakhnov (2022).

Though we do not focus on returns, our empirical findings are related to the finding that central exchange listings tend to increase coin prices (Ante, 2019; van Kampen, 2022; Talamas, 2021; Li et al., 2022). Shams (2020) shows that coins listed on similar exchanges have greater return comovement; this is distinct from, and complementary to, our finding that large exchange listings decrease within-coin price dispersion. Benedetti and Nikbakht (2021) analyze "cross-listings", defined as the second time a coin is ever listed, finding that coin prices, trade volumes, and coin network growth increase after cross-listings. Their analysis is related to, but distinct from, our analysis of the effect of large exchange listings; moreover, Benedetti and Nikbakht do not analyze coin price dispersion, on-chain flows, or listing propensity. We more loosely relate to an empirical literature analyzing exchange fragmentation in equity markets (Foucault and Menkveld, 2008; O'Hara and Ye, 2011), and an empirical literature on OTC markets (Di Maggio, Kermani and Song, 2017; Hagströmer and

Menkveld, 2019; Li and Schürhoff, 2019; Hendershott et al., 2020); we discuss the differences between the crypto and equity exchange settings in Section 2.

We also contribute to a literature analyzing blockchain flow data. Makarov and Schoar (2021) analyze the Bitcoin blockchain, clustering exchanges based on flows, and finding that Binance, Coinbase, and Huobi have highest flow centrality. Cong et al. (2023a) analyze activity concentration on the Ethereum blockchain, though they do not focus on exchanges. Dyhrberg (2020) and Hautsch, Scheuch and Voigt (2024) show that crypto arbitrage flows are associated with exchange price gaps. We believe we are the first to analyze inter-exchange token flows on Ethereum; we complement Makarov and Schoar (2021) in finding that Coinbase and Binance play a central role for Ethereum-based coins.

The paper proceeds as follows. Section 2 describes institutional background around cryptocurrency exchanges. Section 3 describes our model. Section 4 describes our data. Section 5 provides evidence of the core-periphery structure in prices and arbitrage flows, and Section 6 tests our model predictions regarding the effects of central exchange listings. Section 7 presents robustness checks of our empirical results. We discuss our results and conclude in Section 8.

2 Institutional Background

Cryptocurrency exchanges, analogous to exchanges for stocks, bonds, and other financial assets, allow customers to exchange fiat currencies for cryptocurrencies. Crypto exchanges operate as custodial platforms: they allow users to deposit fiat or cryptocurrencies, hold these assets on their behalf, and facilitate trading among users. For the vast majority of exchanges, trading is governed through limit-order books. There are a number of reasons why consumers may purchase cryptocurrencies. Cryptocurrencies may be used as a store of value or medium of exchange, in countries with high inflation or weak financial institutions;¹ market participants also buy cryptocurrencies to speculate, or participate in "decentralized finance" applications.

Traditional financial assets are basically legal constructs: US stocks are generally recorded as electronic entries at the DTCC, and US stocks are generally held by entities with a legal and/or infrastructural presence in the US. Cryptocurrencies differ in that their ownership is instead enforced through cryptographic consensus algorithms. Bitcoin ownership is recorded on the Bitcoin blockchain; digital signature algorithms ensure that only wallet "owners" can

 $^{^1\}mathrm{See}$ CNBC and Rest Of World for a discussion of the use of cryptocurrencies as a store of value in Lebanon.

transfer funds that belong to them. Thus, Bitcoin can be held anywhere individuals can access the internet. There is a single global Bitcoin ledger, implying that Bitcoin in one wallet is fungible with Bitcoin in any other wallet, regardless of where the wallet owners reside.

Cryptocurrency exchanges can be thought of as "on/off-ramps" between global cryptocurrency ledgers and jurisdiction-specific fiat payment systems. Users generally trade on crypto exchanges using exchange-custodied fiat and cryptocurrencies;² however, users can also "withdraw" custodied cryptocurrency into private wallet addresses, or "deposit" crypto by sending cryptocurrency into an exchange to increase their custodied crypto balances. There are a large number of crypto exchanges: according to Blockspot.io, as of 2023 there are over 1,000 active exchanges. This is somewhat puzzling in a very concentrated asset class: the top 2-3 cryptocurrencies account for well over 50% of the entire sector's market cap, according to sources such as Coinmarketcap.

Crypto exchanges differ from the well-studied setting of equity exchanges in several key ways. In traditional equity markets, retail investors almost never interact directly with exchanges; instead, they place orders through brokerages, which may execute internally, route to market makers, or send orders to exchanges. Equity exchanges in many jurisdictions must compete on price due to regulatory requirements: for example, under the SEC's Regulation NMS, brokers must execute trades at the best available price, prohibiting "trade-throughs" where orders are filled at worse prices than publicly available quotes.

In contrast, most crypto exchanges allow retail and institutional investors to trade directly, with minimal intermediation. Services such as fiat transfers, custody, and leverage tend to be handled in-house. Retail investors do not have simple tools to compare prices across exchanges, and there is no regulatory requirement to ensure best execution. Crypto exchanges are also differentiated based on many factors, such as the regulatory jurisdictions they serve, the fiat onramps they support, the extent to which they implement know-your-customer and anti-money-laundering policies, and the extent to which they offer features such as leverage and derivatives trading. These factors suggest that crypto traders may exhibit much higher "exchange stickiness" than equity traders, who (in the US) benefit from forced price competition even if they are indirectly interacting with exchanges.

Yet customers on worldwide crypto exchanges ultimately trade the same few cryptoassets. Price differences across exchanges thus create arbitrage opportunities: if a coin's price on exchange A is lower than its price on exchange B, at current fiat exchange rates, an arbitrageur can buy the coin on A, send the coin to B and sell, and profit after withdrawing fiat balances from B. The literature has documented various costs to inter-exchange arbitrage, such as

²An exception is that some exchanges primarily focus on trading stablecoins for cryptocurrencies.

price risk due to settlement delays (Hautsch, Scheuch and Voigt, 2024); trading, deposit, and withdrawal fees imposed by exchanges; capital controls and FX costs (Makarov and Schoar, 2020; Choi, Lehar and Stauffer, 2022; Dyhrberg, 2020; Borri and Shakhnov, 2023); and exchange-level credit risk. Building on the literature, we give a stylized description of how the inter-exchange arbitrage process incurs these costs in Appendix A. Many of these costs plausibly have large fixed components, implying that specialized arbitrageurs have a comparative advantage over regular investors in exploiting cross-exchange crypto price differences.

We proceed to build a model based on the idea that regulatory and fiat barriers create effective "transportation costs" between groups of traders in different jurisdictions, even when they trade identical assets. As in classical commodity spot markets, we conjecture that this naturally leads to a market structure characterized by small local exchanges within each jurisdiction, connected by specialized arbitrageurs who move inventory across exchanges in response to price differences caused by supply and demand imbalances.

3 Model

We construct a model with a number of ex-ante identical crypto exchanges, with captive customers who receive "inventory shocks" that influence exchange prices for a single risky asset. Arbitrageurs choose pairs of exchanges to link, profiting from price gaps induced by these inventory shocks. We show that, under certain parameter settings, arbitrageurs endogenously coordinate on linking exchanges to a single central exchange, which emerges as a "liquidity hub" in equilibrium.

3.1 Setup

There are N identical exchanges, indexed by i, each populated by a unit measure of fully captive customers. There are N - 1 arbitrageurs, who profit from exploiting price differences across exchanges generated by inventory shocks.³ All agents are risk-neutral.

The model has two stages. In the first stage, each arbitrageur a chooses a pair of exchanges $\{i, j\}$ to connect; arbitrageurs' decisions are simultaneous. In the second stage, inventory

³The assumption that there are exactly N-1 arbitrageurs is artificial but convenient, since this is exactly the number of arbitrageurs needed to form a star network equilibrium. In Appendix E, we construct a simple extension of our model incorporating costly arbitrageur entry, and show that there are entry costs which sustain equilibria where exactly N-1 arbitrageurs enter, and the Nth arbitrageur perceives lower expected profits and thus chooses not to enter.

shocks $x_1 \dots x_N$ are drawn, and exchange markets clear in a simultaneous uniform-price auction.

Customers. Customers on exchange i have utility:

$$u(q_i) = \psi(q_i + x_i) - \frac{\gamma}{2}(q_i + x_i)^2$$
(1)

for purchasing q_i units of the asset. ψ is a constant that represents customers' average value for the cryptoasset; all model outcomes will shift linearly in ψ , so the model's predictions are identical if we simply set $\psi = 0$. x_i is an aggregate inventory shock, common to all customers of exchange i; x_i can be thought of as capturing changes in speculative demand, or increases in economic policy uncertainty (Yu and Zhang, 2022), or other forces which shift average demand across an exchange's customer base. We assume each x_i has mean 0, variance σ , bounded support, and that the x_i are i.i.d. across exchanges. Customers also bear a quadratic inventory cost whose magnitude is increasing in γ .

Customers are atomistic and ignore their price impact. Thus, if the price on exchange i is p_i , we derive customer demand by setting the derivative of (1) equal to p_i and solving:

$$q_i = -x_i - \frac{1}{\gamma} \left(p_i - \psi \right) \tag{2}$$

Intuitively, at price $p_i = \psi$, customers set $q_i = -x_i$, selling exactly as much as they receive in inventory shocks. Net demand is decreasing in prices, and demand is more elastic when γ is lower. Inverting (2), we can express the price on exchange *i* as a function of customers' net purchase quantity:

$$p_i - \psi = -\gamma \left(q_i + x_i \right) \tag{3}$$

We assume exchange i's customers are identical for modelling simplicity; within-exchange trade could be microfounded by assuming the inventory shock x_i has some variance across exchange i's customers.

Arbitrageurs. In the first stage of the model, each arbitrageur a chooses a pair of exchanges $\{i, j\}$ to connect; multiple arbitrageurs are allowed to choose the same pair. In the second stage, an $\{i, j\}$ arbitrageur buys a total amount z_{ij} of coins on exchange i to sell on j. We do not allow arbitrageurs to hold net positions, so the purchase and sale amounts must be equal. Letting p_i and p_j be the relative prices on i and j, arbitrageurs' utility is:

$$V_a\left(z_{ij}^a\right) = z_{ij}^a\left(p_j - p_i\right) - \frac{\zeta}{2}\left(z_{ij}^a\right)^2\tag{4}$$

That is, arbitrageurs incur a "transportation" cost $\frac{\zeta}{2} \left(z_{ij}^a\right)^2$ for moving z_{ij}^a units of the asset, where $\zeta > 0$. This cost captures many frictions to inter-exchange arbitrage documented in the literature, including delays, price risk, and fiat transfer costs (Makarov and Schoar, 2020; Choi, Lehar and Stauffer, 2022; Augustin, Rubtsov and Shin, 2023); we describe some of these costs in Appendix A. In the main text, we assume arbitrageurs behave competitively, ignoring their price impact.⁴ Thus, arbitrageurs increase z_{ij}^a until their marginal transportation cost is equal to the price difference $p_j - p_i$. Setting the derivative of (4) with respect to z_{ij}^a to zero, assuming $p_j - p_i$ is constant, we derive arbitrageurs' supply as a function of the i, j price gap:

$$\zeta z_{ij}^a = p_j - p_i \tag{5}$$

That is, z_{ij}^a increases linearly in the price gap $(p_j - p_i)$, with a greater slope when arbitrageurs' transportation costs ζ are lower. Plugging (5) into (4), arbitrageur profits are proportional to the squared final price difference between exchanges:

$$V_a(z_{ij}) = \frac{1}{\zeta} (p_j - p_i)^2 - \frac{1}{2\zeta} (p_j - p_i)^2 = \frac{1}{2\zeta} (p_j - p_i)^2$$
(6)

Market Clearing. In the second stage, all exchanges clear in a simultaneous uniformprice double auction. Arbitrageurs' and customers' demand must sum to zero for each exchange, implying:

$$q_i = \sum_{a \in \mathcal{A}_i} z_{ji}^a \tag{7}$$

where \mathcal{A}_i is the set of all arbitrageurs *a* who connect to exchange *i*. Using (7) to replace q_i in customers' optimality condition (3), we can define equilibrium in our model as a set of exchange prices p_i and arbitrage flows z_{ij}^a which satisfy customers' and arbitrageurs' demand equations respectively.

Definition 1. An equilibrium, given inventory shocks x_i , is described by a set of prices p_i for all i, and arbitrage flows z_{ij}^a for each arbitrage a, satisfying customer optimality for each exchange i:

$$p_i - \psi = -\gamma \left(\sum_{a \in \mathcal{A}_i} z_{ji}^a + x_i \right) \tag{8}$$

and arbitrage optimality for each arbitrage a:

$$\zeta z_{ij}^a = p_j - p_i \tag{9}$$

⁴The competitive-bidding assumption is strong, but substantially simplifies the model. In Appendix D, we extend the model to include strategic bidding, and show that star-network equilibria continue to exist, though under a slightly stricter cutoff on the parameters ζ , γ .

We can rearrange (8) and (9) slightly to:

$$-\frac{p_i - \psi}{\gamma} - \sum_{a \in \mathcal{A}_i} z_{ji}^a = \qquad \qquad x_i \tag{10}$$

$$\zeta z_{ij}^a - (p_j - p_i) = 0 (11)$$

The equilibrium conditions can then be written as the matrix equation:

$$\boldsymbol{M}\begin{pmatrix} p_{1}-\psi\\ \vdots\\ p_{N}-\psi\\ z^{1}\\ \vdots\\ z^{N-1} \end{pmatrix} = \begin{pmatrix} x_{1}\\ \vdots\\ x_{N}\\ 0\\ \vdots\\ 0 \end{pmatrix}$$
(12)

with \boldsymbol{M} representing the LHS of the set of equations (10) and (11). The system (12) is explicitly solvable by inverting \boldsymbol{M} . The resultant outputs z^a and $p_i - \psi$ are linear in the inventory shocks $x_1 \dots x_N$; this immediately implies they have mean 0, and variance and covariances which are a function of \boldsymbol{M} and σ . The expectations of any quadratic function of the outputs, such as expected squared trade volumes $(z^a)^2$, or expected squared price differences $(p_i - p_j)^2$ – which determines arbitrageur profits from (6) – are similarly simple functions of \boldsymbol{M} and σ .

Figure 1 illustrates equilibrium outcomes for two example networks: we derive analytical solutions for both cases in Appendix B.1 and B.2. The top panel illustrates two connected exchanges, with $x_1 = 2$ and $x_2 = -1$. With frictionless flows, final inventories and prices should be equalized: we should have $z_{12} = 1.5$, so inventory is 0.5 and prices are -1 on each exchange. With frictions, arbitrage flows move partially in this direction: we have $z_{12} = 1$, so exchange 1 is left with slightly higher inventory and lower prices than exchange 2. Our model thus captures the empirically documented fact that crypto market arbitrageurs trade in the direction of price gaps (Dyhrberg, 2020; Makarov and Schoar, 2021; Hautsch, Scheuch and Voigt, 2024), but do not fully close these gaps (Kroeger and Sarkar, 2017; Makarov and Schoar, 2020; Yu and Zhang, 2022; Choi, Lehar and Stauffer, 2022; Borri and Shakhnov, 2023).

The bottom panel shows a star network: four exchanges $1 \dots 4$ are connected to a center C. Analogous to the two-exchange case, arbitrageurs move inventory from exchanges 1 and 2 towards 3 and 4, but arbitrage flows imperfectly equalize initial price differences. In this case,

the central exchange – despite being fundamentally identical to other exchanges – behaves as a trading hub: *net* trade quantity is low, since the initial and final inventory positions are similar, but *gross* trade quantity is high, since peripheral exchanges trade through their arbitrage links to the central exchange.

Our model has a convenient analogy to a spring system, where exchanges are nodes and arbitrageurs are springs. Spring rigidity ζ controls price dispersion. As $\zeta \to 0$, springs are fully rigid and prices align perfectly across exchanges; for $\zeta > 0$, shocks propagate imperfectly, and nodes further from the initial shock are influenced less. As in spring systems, the effects of inventory shocks to different nodes combine linearly.

3.2 Results

Proposition 1. When

$$\frac{\zeta}{\gamma} < 2.21432\dots \tag{13}$$

when N is sufficiently large, there exists a star-network equilibrium, in which each arbitrageur connects an exchange to a common central exchange.

Proof. Appendix C contains the full proof; we present a sketch here. The conditions for star equilibria to exist are depicted in Figure 2. Suppose N - 2 arbitrageurs form a star, leaving one isolated exchange unconnected to the center. The last arbitrageur has four choices: she can join another arbitrageur on a central-peripheral link (red), link the isolated exchange to another peripheral exchange (green), link two peripheral exchanges (orange); or complete the star by linking the isolated exchange to the central exchange (blue). Star-shaped equilibria exist if blue is optimal: that is, if N - 2 arbitrageurs are expected to form a star, the final arbitrageur optimally completes the star.

The case for finite N is in principle solvable but complex, so we instead analyze the limit as $N \to \infty$. In the limit, the center node of the (N-2)-node star converges to having fixed price ψ , and infinite market depth, since it is connected to increasingly many peripheral nodes. It is thus straightforward to characterize the limiting payoffs of the four strategies. We find that blue dominates red, green dominates orange, and blue is preferred to green in the limit depending on (13). This implies that blue will be optimal for sufficiently high N whenever (13) holds, proving the proposition.

The intuition behind Proposition 1 is that arbitrageurs face a tradeoff between initial price differences and price impact: two peripheral exchanges (green) will tend to have larger initial price differences than a peripheral exchange and the central exchange (blue), but the central exchange is deeper than peripheral exchanges. When ζ is high relative to γ , arbitrageurs are capacity-constrained due to their own transportation costs; they thus care less about price impact, and prefer connecting peripheral exchanges with large price differences. When ζ is low, arbitrageurs trade larger amounts and care more about price impact, leading them to prefer connecting to the deep liquidity available at the central exchange.

Proposition 1 parallels the classic result that liquidity drives traders to concentrate on a single trading venue in equilibrium (Admati and Pfleiderer, 1988; Pagano, 1989). However, whereas classic models analyze *traders*' decisions, we analyze *arbitrageurs*' centralization incentives. Our conclusion is nuanced: there exist parameter settings under which arbitrageurs prefer connecting *less* liquid exchanges, so no centralized equilibrium exists.

We note that Proposition 1 demonstrates existence, but not uniqueness, of star-network equilibria; even when (13) holds, we have not found a way to rule out the existence of non-star-shaped equilibria.

3.3 Implications

Prediction 1. Price correlations follow a star structure: peripheral exchange prices are more highly correlated with central exchange prices than with those of other peripheral exchanges.

Prediction 2. Arbitrage flows across exchanges have a core-periphery structure: flows tend to link "peripheral" exchanges to one or a few "central" exchanges.

These two predictions show that the star structure of equilibria should be visible in both prices and quantities. Formally, Prediction 1 can be seen clearly in the limit as $N \to \infty$: central exchange prices are ψ , and peripheral exchange prices are $\psi + \epsilon_i$, so peripheral prices are equal to the central exchange price plus i.i.d. noise. Intuitively, central exchange prices are insensitive to its own inventory shocks, as well as shocks to any particular peripheral exchange, because inventory can efficiently flow from the central exchange to all peripheral exchanges. Peripheral exchange prices are "noisier", because each peripheral exchange only has a singular link to the center. Prediction 2 is immediate in our model: arbitrageurs move inventory across ex-ante identical exchanges by coordinating on one exchange as a "hub", so arbitrage flows should be larger at the central exchange.

3.4 Listing Decisions

Next, we analyze the effects of central exchange listings on market outcomes. Suppose arbitrageurs have formed a star-shaped network, and consider peripheral exchanges' incentives to list a second risky coin, which has inventory shocks on exchanges identical to the baseline model. Assume total trade volume for the new coin on exchange *i* consists of arbitrage volume q_i , plus some exogeneous trade volume V_i . Peripheral exchanges have some volume threshold M_i above which they are willing to list the coin.⁵ We suppose N is large, and that some large number of peripheral exchanges have already listed the coin.

Prediction 3. Coin price dispersion across peripheral exchanges decreases when central exchanges list.

Formally, the dispersion of prices across peripheral exchanges, Var(p), is $\gamma^2 \sigma^2$ when peripheral exchanges are in autarky; using (62) from Appendix C.3.1, dispersion decreases to $\gamma^2 \left(\frac{\zeta}{\gamma+\zeta}\right)^2 \sigma^2$ when peripheral exchanges are connected to the infinitely deep central exchange. Intuitively, in our model, peripheral exchanges are connected through the central exchange: only after the central exchange lists the coin can arbitrageurs trade to smooth out the effects of inventory shocks.

A large empirical literature has documented that crypto prices vary substantially across peripheral exchanges, in a way that appears to be associated with various measures of arbitrage frictions, such as currency and geographical boundaries (Makarov and Schoar, 2020; Dyhrberg, 2020; Choi, Lehar and Stauffer, 2022; Borri and Shakhnov, 2023), settlement delays (Hautsch, Scheuch and Voigt, 2024), and the availability of derivatives (Augustin, Rubtsov and Shin, 2023). Our model suggests another factor that should influence price dispersion: if large exchanges indeed behave as arbitrage "hubs", then coin price dispersion across small exchanges should decrease after large exchanges list a coin. If, instead, large exchanges simply have large customer bases, and small exchanges are mostly arbitraged against each other, large exchange listings should not substantially impact small exchange price dispersion.

Prediction 4. When the central exchange lists a coin, peripheral exchange arbitrage flow volumes tend to increase.

In our model, price dispersion decreases after central exchange listings because inventory shocks result in *quantity* flows to the central exchange, alleviating the effects of inventory shocks on prices; we should thus also observe increases in gross on-chain arbitrage flows from peripheral exchanges after central exchanges list.

Prediction 5. When the central exchange lists a coin, peripheral exchange total trade volumes tend to increase.

⁵This can be microfounded, for example, by assuming peripheral exchanges have some fixed cost of listing, and charge some trading fee τ_i per unit volume, where τ_i is small enough that we can ignore its effect on market outcomes.

Prediction 6. Peripheral exchanges are more likely to list a given coin after the central exchange lists the coin.

In our model, total trade volume is simply exogeneous volume V_i plus customer-toarbitrageur trade volume q_i . When central exchanges list, we should thus observe increases in total trade volume. Anticipating increased volumes, peripheral exchanges which had not already listed the coin may then decide to list.

These predictions are straightforward within our model. However, they contrast with the predictions of an alternative theory where large exchanges emerge because of *customers*' tendency to seek liquidity rather than arbitrageurs. In such a model, large exchange listings would *decrease* small exchanges' trade volumes and listing probabilities, as customers flock to trade in the more liquid large exchanges.

3.5 Discussion of Model Assumptions

Our model can be thought of as a demand-schedule submission game (Rostek and Yoon, 2020), with a few key simplifications relative to the literature. First, we fully separate the roles of customers, who can hold inventory, and arbitrageurs, who can transport inventory across exchanges but cannot hold net positions. Second, the baseline model assumes competitive bidding; Appendix D analyzes the case with price impact. Third, we impose strict participation constraints: customers are captive to exchanges, and each arbitrageur trades on a single pair of exchanges. These assumptions are stylized, but lead to a very simple model of frictional arbitrage between exchanges. The assumption that there are exactly N - 1 arbitrageurs is similarly artificial; we discuss relaxing it in a simple costly-entry model in Appendix E.

The assumption that each arbitrageur must commit to a single pair of exchanges captures the idea that integrating with an exchange's trading infrastructure, regulatory restrictions, and fiat on- and off-ramps is costly. In practice, arbitrageurs may trade on more than two exchanges at once. A model where arbitrageurs choose sets of exchanges to participate on would be much more complex – transportation costs and trading quantities would be more complicated, and it is not even clear what a core-periphery network would look like in this setting – so we assume arbitrageurs can only choose pairs of exchanges for simplicity.

Strategic complementarities in the model imply that, even if one exchange were slightly larger $(x_i \uparrow)$ or deeper $(\gamma_i \downarrow)$ than others, it may not emerge as central, if arbitrageurs coordinate on a smaller exchange as a hub. This introduces unhelpful indeterminacy in the model, so we assume exchanges are homogeneous for simplicity. We similarly assume homogeneous ζ for simplicity, though in practice arbitrage costs likely vary for different pairs of exchanges, as shown by Dyhrberg (2020).

We assume captive customers; our primary motivation for this is fiat and regulatory boundaries facing customers, but exchanges could also be differentiated based on factors such as search frictions and advertising; differentiation in exchanges' product offerings, such as the availability of leverage and derivatives trading; or the degree to which exchanges implement know-your-customer and anti-money-laundering policies, in an attempt to avoid interacting with sanctioned individuals or funds. It is not important for the model what the drivers of exchange differentiation are, so long as they imply that each exchange's customers have difficulty switching to other exchanges.

Exchange customers are of course not perfectly captive in practice. If all customers could frictionlessly move across exchanges, classic results suggest that trade will concentrate on a single exchange (Pagano, 1989), counter to what we observe in crypto markets empirically. If instead we assumed that some fraction of customers can switch exchanges, Predictions 5 and 6 could change sign: central exchanges could cannibalize peripheral exchange volumes. Incorporating partially elastic customers complicates the model, and we think it is unlikely to produce other nonobvious predictions, so we assume full captivity for simplicity.

Our model does not feature the "listing pump" effect, that central exchange listings tend to associate with increased coin prices (Ante, 2019; van Kampen, 2022; Talamas, 2021). This could be microfounded either through a "liquidity premium" effect, or through a number of other channels we discuss in Section 7.

We assume there is a single central exchange for tractability; in our empirical analysis, we treat Binance and Coinbase as central exchanges, and the long tail of smaller exchanges as peripheral.

4 Data

4.1 Cryptotick Data

The primary dataset used in this paper is from cryptotick.com, which collects coin trade price and quantity data from a broad set of cryptocurrency exchanges. Cryptotick obtains this data by querying APIs provided by the exchanges and timestamps the data using a synchronized clock (UTC time) for all exchanges. The dataset contains hourly OHLCV data—that is, open, high, low, and close prices—as well as total trade volume for each hour on each exchange. Each data series in the Cryptotick dataset represents a trading pair on a given exchange. A trading pair consists of one cryptocurrency traded against either a fiat currency, a stablecoin (a cryptocurrency designed to maintain a fixed value relative to a fiat currency), or another major cryptocurrency such as BTC or ETH. For our analysis, we aggregate the data to a daily frequency for each trading pair-exchange ID. Specifically, we compute the daily average price as a volume-weighted mean of hourly open prices and sum the trade volumes across all hours within a day.

Our dataset spans from January 2017 to December 2023 and includes 267 exchanges and over 10,000 coins in the raw dataset. We restrict our sample in two ways: by the cryptocurrency in the trading pair, and by the denominator that the cryptocurrency is traded against. Since many coins are not actively traded, we first limit our sample to trading pairs that involve the top 200 cryptocurrencies ranked by coinmarketcap.com on June 1 in any year from 2017 to 2023. We further restrict our sample to three types of trading pairs: pairs involving one of 27 major fiat currencies;⁶ pairs involving BTC or ETH, the two largest cryptocurrencies by market capitalization; and pairs involving one of the three major stablecoins (USDT, USDC, and BUSD).

For all trading pairs, we convert coin prices and volumes to USD terms. For fiat pairs, we use same-day USD-to-fiat exchange rates, and for crypto pairs, we use daily prices of cryptocurrencies and stablecoins from Yahoo Finance. We aggregate data to coin-level by taking the average of the prices for each trading pair involving same coin within a day weighted by its USD trade volume, and adding volumes across all trading pairs of the same coin within a day. After dropping stablecoins and fiat currencies in the final step, the final sample consists of 770 coins across 254 exchanges. Appendix F.1 contains more details about the data cleaning process.

We identify the listing date of a coin on an exchange by observing the first date it appears on an exchange in our price and volume data. We round to the nearest day: if the first trade we observe occurs before 12:00PM, we identify the listing date as the previous day, and if the first trade occurs after 12:00PM, we identify the listing date as the current day. Summary statistics of the data are shown in Panels A to C of Table 1.

4.2 On-Chain Data

We collect data on token deposits and withdrawals from exchanges on the Ethereum blockchain, as most coins in our sample are ERC-20 tokens, and the majority of stablecoin transactions occur on Ethereum. Using Etherscan, we manually identify exchange-owned wallet addresses. Etherscan labels wallets associated with exchanges and other entities, allowing us to search

 $^{^6{\}rm These}$ 27 major fiat currencies are: NZD USD KRW JPY CNY IDR SGD VND TWD AUD PKR ZAR TRY MXN BRL CHF ILS PLN GBP RUB EUR CAD HKD INR SAR AED SEK.

for exchange names in our dataset and apply filters detailed in Appendix F.2.1. We then collect all transactions associated with these wallet addresses. For comparability across coins, we dollarize all arbitrage flows, multiplying by a daily average price for each coin, calculated as the volume-weighted average of USD prices in the Cryptotick data.

We attempt to identify sequences of Ethereum transactions that likely represent "arbitrage trades" across exchanges. Mechanically, on Ethereum, exchanges typically maintain "main" wallets holding most balances and trader-specific "deposit addresses", whose private keys are controlled by the exchange, which allow them to distinguish deposits sent by different traders. Traders can then "deposit" into the exchange simply by sending crypto to their deposit address; the exchange then credits the trader's custodial balance of the token, generally maintained in an internal non-blockchain database at the exchange, with the deposited tokens, and over time generally "sweeps" the deposited crypto into a "main" exchange wallet. A trader "withdraws" by sending a request to the exchange to send custodied crypto to a specified wallet address; the exchange then decreases the trader's custodial balance, and sends crypto from the "main" exchange wallet to the specified wallet.

We define arbitrage transactions as withdrawals at one exchange, paired with deposits at another exchange. We found that this generally occurs through traders directly ordering withdrawals from one exchange into an unlabelled wallet that we infer to be a deposit address at another exchange; these tokens are later "swept" into a labelled exchange wallet address. We thus define arbitrage transactions somewhat restrictively, as withdrawals from a given exchange wallet, whose balances are entirely converted into deposits within 48 hours and with transaction amounts ranging from 99% to 101% of the withdrawal amount. Appendix F.2.2 describes further details of this process.

Although restrictive, our criteria classify approximately 8% of all deposits and withdrawals from exchanges as arbitrage transactions. Overall, our procedure identifies around 21 million transactions across 239 coins and 93 exchanges. Summary statistics of the data are shown in Panel D of Table 1.

5 Core-Periphery Structure in Prices and Arbitrage Flows

5.1 Core-Periphery Structure of Price Correlations

Figure 3 plots the CDF of Bitcoin return correlations, separately for peripheral exchange pairs, peripheral-Binance pairs, and peripheral-Coinbase pairs. Consistent with Prediction 1, the red line lies above the blue and green lines: peripheral-central return correlations tend to be higher than peripheral-peripheral correlations. Quantitatively, the 25th, 50th, and 75th percentiles of return correlations among all exchanges are 0.59, 0.78, and 0.91. The percentiles are 0.76, 0.86, and 0.94 for pairs involving Binance, and 0.65, 0.83, and 0.94 for pairs involving Coinbase.⁷

While our model is static, a simple dynamic interpretation comes from assuming that arbitrage occurs with a lag, so exchange prices initially reflect inventory shocks x_i , and move towards reflecting the net quantity $x_i + q_i$ over time as arbitrageurs trade. The example in the bottom panel of Figure 1 then suggests that peripheral-central price gaps should close mainly through peripheral exchange prices adjusting: arbitrage quantity flowing into the central exchange has lower price impact, since it is spread out over all other peripheral exchanges.

To test these hypotheses, we define the log price gap between peripheral exchange e and a central exchange, for coin c at time t, as:

$$PriceGap_{c,e,t} \equiv p_{c,e,t} - p_{c,t}^{cert}$$

That is, $PriceGap_{c,e,t}$ is the difference in log prices between e and the central exchange for coin c at time t. A number of papers show that cross-exchange price deviations are mean-reverting (Kroeger and Sarkar, 2017; Makarov and Schoar, 2020; Dyhrberg, 2020; Borri and Shakhnov, 2023); we show this holds also in our sample, by estimating the specification:

$$\Delta PriceGap_{c,e,t} = \beta_{c,e}^{PriceGap} PriceGap_{c,e,t-1} + \epsilon_{c,e,t}^{PriceGap}$$
(14)

⁷The average return correlation across all exchanges at day level in our dataset is 0.7. However, the low cross-correlation is largely driven by low ranking exchanges in our representative dataset. Our data produces price correlations very close to those in the literature, once we subset to the same set of exchanges. Makarov and Schoar (2020) use data from Kaiko, and restrict their sample to 17 large exchanges; they measure an average hourly cross-correlation of USD BTC returns of 0.83. After restricting our sample to the same set of exchanges, we find a very similar average cross-correlation of 0.83. Return correlations are sensitive to the time horizon at which prices are measured: at the daily level, Makarov and Schoar report a higher average cross-correlation of 0.95.

We estimate $\beta_{c,e}^{PriceGap}$ for each coin-exchange pair. The distribution of estimated $\beta_{c,e}^{PriceGap}$ coefficients is shown in the top row of Figure 4. The estimated coefficients concentrate at -1. The median coefficient is -0.82 (-0.89), implying that a 10% price gap on average decays to 1.8% (1.1%) on average within a day treating Binance (Coinbase) as the central exchange; in other words, it takes roughly 1.35 (1.05) days for a price gap to tenth. The 75th percentile $\beta_{c,e}^{PriceGap}$ coefficient implies a price gap tenth-life of 2.81 (2.03) days, and the 25th percentile implies a tenth-life of 0.78 (0.57) days, again for Binance (Coinbase) as the central exchange. In other words, these figures show that price gaps revert towards 0 fairly quickly, rarely lasting more than a few days.

We then analyze the direction of price convergence, by regressing peripheral and central exchange log returns on lagged price gaps:

$$\Delta p_{c,e,t} = \beta_{c,e}^p PriceGap_{c,e,t-1} + \epsilon_{c,e,t}^p \tag{15}$$

$$\Delta p_{c,t}^{cen} = \beta_{c,e}^{p^{cen}} PriceGap_{c,e,t-1} + \epsilon_{c,e,t}^{p^{cen}}$$
(16)

The coefficient estimates are respectively shown in the middle and bottom rows of Figure 4. The results strongly support the prediction. The middle row shows that the distribution of coefficients from Specification (15) is concentrated at negative values: when $PriceGap_{c,e,t-1}$ is positive, peripheral exchange prices tend to decrease towards central exchange prices. Quantitatively, taking the median coefficient, a 10% price gap predicts a 6% (9.9%) decrease in return on peripheral exchanges on the following day, treating Binance (Coinbase) as the central exchange, respectively.

In contrast, the distribution of estimated $\beta_{c,e}^{p^{cen}}$'s is roughly symmetric around 0 with median value of 0.01 (-0.13) when treating Binance (Coinbase) as the central exchange; thus, $PriceGap_{c,e,t-1}$ does not positively or negatively predict returns on central exchanges.

We also estimate a single coefficient for $\beta^{PriceGap}$, β^p , and $\beta^{p^{cen}}$ and show the results in Appendix Table A.1. Given dynamic and cross-sectionally dependent panel data, the OLS estimator is biased and inconsistent. Following Pesaran (2006), we use a mean group estimator, which is consistent when both N and T are large—valid in our data. We reach a similar conclusion.

5.2 Core-Periphery Structure of Arbitrage Flows

The literature has documented that Bitcoin flows tend to have a core-periphery structure, clustering around a few central exchanges (Makarov and Schoar, 2021); we find analogous

results using our data on Ethereum coin flows, supporting Prediction 2 of our model.

We first aggregate dollarized on-chain arbitrage transactions, at the exchange-pair and exchange levels, across our entire sample period; we call these aggregates respectively $ArbVol_{e,\tilde{e}}$ and $ArbVol_e$. In Table 2, we show the market shares of total dollarized arbitrage and trade volumes, for the top 10 exchanges in terms of arbitrage volume. Consistent with Makarov and Schoar (2021)'s findings on Bitcoin on-chain flows, Binance and Coinbase emerge as large exchanges in our data, in terms of both arbitrage and trade volume.⁸

In our model, the central exchange is large because of the coordination of *arbitrageurs*. This contrasts with two other hypotheses. First, large exchanges may just have fundamentally larger customer bases than small exchanges. Second, *customers* could coordinate on large exchanges, due to their greater liquidity. If indeed arbitrage is an important contributor to centrality, we would expect arbitrage volumes on central exchanges to be large *relative to* trade volumes.⁹ This seems to hold empirically: Binance's share of total arbitrage volume is slightly larger than its share of trade volume, and Coinbase's share of arbitrage volume is more than four times its share of trade volume. This provides some support for our model's view that central exchanges serve as *arbitrage* hubs, not just centers of trading liquidity.

We then convert the exchange-pair arbitrage flow data into an unweighted graph of "arbitrage links" between exchanges, in order to visualize the arbitrage network, as well as to construct network centrality measures for different exchanges. We assume two exchanges e and \tilde{e} are "connected" whenever the arbitrage flows between e and \tilde{e} are a large fraction of total arbitrage flows involving either exchange: formally, if $ArbVol_{e,\tilde{e}}$, is at least 10% of the smaller of $ArbVol_e$ and $ArbVol_{\tilde{e}}$. Using the resultant network of links between exchanges, we calculate three measures of the "network centrality" of each exchange: degree centrality, betweenness centrality, and eigenvector centrality; we define these measures formally in Appendix G.

Appendix Figure A.1 visualizes the resultant network of exchanges; the size of each node is linked to the number of edges connected to the node. Coinbase and Binance emerge visually as large exchanges; other exchanges that are fairly large include Huobi, Bitfinex, FTX, and Kraken. Our stylized model makes the sharp prediction that arbitrage networks should have a star structure, with peripheral exchanges linked only to a single central hub: while the true network is obviously more complex, we do observe a number of small exchanges that have a

⁸FTX has slightly higher total arbitrage volume than Coinbase in our data. We do not treat FTX as a central exchange in our analysis because it failed during our sample period; we are thus not fully confident in its data around the time of its failure, and we are also not fully confident that on-chain flows to and from FTX reflect legitimate arbitrage flows, versus for example proprietary withdrawals.

⁹This holds formally in our model: as $N \to \infty$, arbitrage volume through central exchanges becomes infinite, so trade volume becomes dominated by arbitrage flows in the limit.

single link to either Binance or Coinbase. Table 2 shows our three centrality measures for each large exchange: Binance emerges as central under all centrality measures, and Coinbase tends to rank highly among exchanges under most measures.

6 Central Exchange Listings and Peripheral Exchange Outcomes

Next, we test Predictions 3 to 6 of the model, regarding the effects of central exchange listings on price dispersion, arbitrage and trade volumes, and peripheral exchange listing propensities. We test these predictions by running coin and coin-exchange level difference-in-differences specifications around central exchange listing events:

$$Y_{c,t} = \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t}$$
(17)

$$Y_{c,e,t} = \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$
(18)

where the variables $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are indicators for whether a coin was listed within the past 30 days or more than 30 days ago, respectively. $PreThreedayListing_{c,t}$ is an indicator equal to 1 for the three days prior to listing; we include this because Binance and Coinbase typically announce listings 1-2 days in advance, so this indicator throws out any pre-listing announcement effects. δ_c and η_t in (17) are coin and time fixed effects, and $\delta_{c,e}$, η_t , and $\gamma_{e,t}$ in (18) are coin-exchange, time, and exchange-time fixed effects. Standard errors are clustered at the coin and time levels for coin level regressions, and at the coin-exchange pair and time levels for coin-exchange level regressions.

We show the results from flexible event-study specifications in appendix figures:

$$Y_{c,t} = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$
(19)

$$Y_{c,e,t} = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$
(20)

where $treat_{c,k,t}$ is a series of dummies indicating days relative to the listing event, and $treat_{c,-31,t}$ and $treat_{c,31,t}$ are dummy variables that equal 1 when time t is more than 30 days before and after the central exchange listing date, respectively. We normalize β_{-30} to 0, so observations 30 days before listing serve as the reference group. We plot estimated coefficients

of $treat_k$ with k from -30 to 30.

6.1 Price Dispersion

Figure 5 plots coin price dispersion around central exchange listing dates. For each listed coin, we calculate price dispersion within a group of "incumbent" exchanges which list the coin at least 30 days before the central exchange lists; thus, we ignore entrants which list shortly before or after central exchange listings. Price dispersion across small exchanges in our sample is substantial: the one month average pre-listing p75-p25 spread is 600bps for Binance listed coins and 400bps for Coinbase listed coins. Consistent with Prediction 3, dispersion declines substantially after listings: in the 30 days after listing, the p75-p25 spread declines by 130bps and 130bps for coins listed by Binance and Coinbase respectively.

We then estimate coin level DID specifications using (17). The dependent variable is coin-day level price dispersion, calculated as the standard deviation of log prices across exchanges. For treated coins, we continue to calculate dispersion using only incumbent exchanges. We also include "never-treated" coins in our sample as is common practice in the DID literature, and for never-treated coins we simply calculate price dispersion using all exchanges in our sample. The DID estimates are shown in Columns (1) and (2) of Table 3. Central exchange listings appear to have a statistically and economically significant effect on price dispersion. Short-run effects for Binance are approximately -0.039, or 32% of initial dispersion, and -0.04 (33%) for Coinbase. Long-run effects are slightly larger, at -0.086 (72%) for Binance and -0.064 (54%) for Coinbase. Appendix Figure A.2 shows results from the flexible event-study specification; pre-listing coefficients are statistically insignificant, so there is no evidence that the parallel trends assumption is violated.

A large literature has analyzed determinants of cross-exchange cryptocurrency price dispersion (Kroeger and Sarkar, 2017; Makarov and Schoar, 2019, 2020; Dyhrberg, 2020; Tsang and Yang, 2020; Borri and Shakhnov, 2022; Yu and Zhang, 2022; Choi, Lehar and Stauffer, 2022; Borri and Shakhnov, 2023; Hautsch, Scheuch and Voigt, 2024). We believe we are the first to show that central exchange listings tend to decrease price dispersion; this finding is possible due to our large panel of exchanges, and focus on a large number of comparatively small coins.

6.2 Arbitrage Volumes

We construct a coin-day level total dollarized arbitrage flow dataset as follows. For coins listed by Binance or Coinbase, we consider arbitrage flows that involve the central exchange, and "incumbent" exchanges which list the coin at least 30 days before the central exchange lists. We do this to avoid incorporating mechanically increased arbitrage flows from exchanges that list the coin after, or shortly before, Binance and Coinbase list. For "never-treated" coins, which are never listed by a central exchange, we simply add all arbitrage flows we observe. Intuitively, the coin-level DID specification is identified through unusual trends in arbitrage volumes involving a fixed set of "incumbent" exchanges and the central exchange, after the central exchange lists, where time fixed effects are identified by coins listed at other dates, as well as "never-treated" coins.

In the resultant data, 40% of coin-day level arbitrage flows are equal to 0, so any effects we estimate will combine "extensive margin" effects of changing the probability of nonzero arbitrage flows, and "intensive margin" effects of changing the size of arbitrage flows conditional on being positive. Chen and Roth (2024) argue that it is difficult to estimate percentage effects in settings where zeros are prevalent, and that the common approach of using $\log (1 + Y)$ as a dependent variable has problematic properties. Following their suggestions, we separately estimate the extensive and intensive margins. Specifically, we use a dummy variable indicating whether arbitrage flows for the unbalanced panel of coin-day pairs with positive arbitrage flows for the intensive margin. In Appendix H.2, we show that our results are robust to using the $\log (a + Y)$ for various choices of the constant a, and running regressions on a subsample that includes only coins where the dependent variables are all greater than zero during the two-month window around central exchange listings.

Columns (3) and (4) in Table 3 present the extensive margin results. We find that central exchange listings significantly increase the probability that arbitrage volumes are nonzero, by 41pp and 25pp for Binance and Coinbase listings respectively. Columns (5) and (6) present the intensive margin results, where we take the logarithm of dollarized arbitrage volumes, and exclude zero-observation days. On-chain arbitrage volume conditional on being positive increases by 220% and 200% for Binance and Coinbase listings in the short term, and by 190% and 170% in the long term, respectively. Appendix Figure A.3 and A.4 show estimates from the flexible event-study specifications; we find no evidence that the parallel trends assumption is violated in any case.

6.3 Trade Volumes

Coin-exchange-day level trade volumes are equal to 0 for 20% of observations: this is somewhat lower than the case of coin-day arbitrage volumes, but still high, so we continue to estimate intensive and extensive margins separately. Similarly, we only keep "incumbent" exchanges for coins listed by central exchanges, but we keep all exchanges for "never-treated" coins. Table 4 presents estimates from Specification (18). The eight columns vary by intensive/extensive margin, fixed effects, and central exchange listings. The first four columns analyze extensive margin, while the last four focus on intensive margin. Columns (1), (2), (5), and (6) include day and coin-exchange fixed effects, while Columns (3), (4), (7), and (8) add exchange-time interactions to account for unobserved heterogeneity at the exchange level.

We find that Listing(0-30 days) coefficients are positive and significant across all specifications. Columns (1) to (4) shows that Binance and Coinbase listings are associated with significantly higher probabilities that trade volumes are nonzero. Columns (5) to (8) show that Binance listings are associated with trade volume increases, conditional on trade volumes being positive, within 30 days by 74% to 81%, while Coinbase listings increase volumes by 77% to 78%. These effects appear to persist beyond 30 days.

Appendix Figures A.5 and A.6 present the results from the flexible event-study specification. The difference between the treated and control groups is small in magnitude prior to listings, though there is a slight pre-trend prior to 3 days before listings, and a significant increase in volumes for Coinbase in the 3 days before listings. Log trade volumes on peripheral exchanges rise sharply following central exchange listings, then gradually decline but remain positive and significant for up to 30 days post-listing.

6.4 Listing Following

Figure 6 plots a simple histogram of peripheral exchange listing times relative to central exchange listings, including only coins that were listed on any exchange for at least 30 days before being listed on central exchanges. Consistent with Prediction 6, there is clear clustering: a large number of peripheral exchanges list just after the central exchange listing.

We then run coin-day DIDs, using the net number of new listings of the coin on each day as the dependent variable. We keep coins that were listed on any exchange for at least 30 days before being listed on central exchanges, or were not listed by central exchanges, eliminating the mechanical effects associated with coins initially listed on central exchanges. Estimates are shown in Columns (7) and (8) of Table 3. Daily peripheral exchange listing rates increase by 0.079 and 0.077 after Binance and Coinbase listings, respectively, translating to an increase of 2.4 and 2.3 exchanges listing the coin in the first 30 days. The long-run effects are economically small, though positive for Binance and negative for Coinbase. Appendix Figure A.7 shows that there is no evidence for pre-trends. These results are intuitively consistent with our model, which suggests that central exchange listings should be quickly followed by

listings on peripheral exchanges, but should not lead to a sustained increase in peripheral exchange listing rates in the long run.

6.5 Interaction between Price Correlations, Volume Increases, and Listing Following

Our model takes the stance that the effects of central exchange listings on price correlations, volumes, and listing increases are all driven by peripheral-central arbitrage. If this is the case, then peripheral exchanges that rely more on central exchanges for arbitrage should tend to have larger volume increases after central exchange listings, and greater tendencies to follow central exchange listings. Our model suggests that a simple measure of the importance of the arbitrage links between a peripheral and central exchanges is price correlations: if a peripheral exchange's price is more correlated with the central exchange, arbitrage forces are likely to be stronger. Thus, we now show that the "volume pump" and "listing following" effects are stronger for peripheral exchanges which have higher price correlations with the central exchange.

6.5.1 Correlation Structure and Volume Increases

To examine the relationship between peripheral-central price correlations and "volume pump" effects, we estimate the following DID specification:

$$Y_{c,e,t} = \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(0-30 \text{ days})_{c,t} \times Correlation_e + \beta_3 Listing(> 30 \text{ days})_{c,t} + \beta_4 Listing(> 30 \text{ days})_{c,t} \times Correlation_e + \beta_5 PreThreedayListing_{c,t} + \beta_6 PreThreedayListing_{c,t} \times Correlation_e + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$

$$(21)$$

where $Correlation_e$ measures the BTC return correlation between exchange e and a central exchange over the full sample period. The results in Table 5 show that the coefficients β_2 and β_4 for the intensive margin are mostly positive and significant, confirming that highcorrelation exchanges experience larger volume increases. In our preferred specifications for trade volume (Columns (7) and (8)), a 0.01 increase in BTC return correlation with Binance (Coinbase) is associated with an additional 1.9% (0.43%) volume increase in the short term, and a 3.4% increase (0.52% decrease) in the long term. The coefficients β_2 and β_4 for the extensive margin are generally negative, suggesting that the effect is mostly on the intensive margin.

6.5.2 Correlation Structure and Listing Following

We then test whether peripheral exchanges with higher price correlations to central exchanges are also more likely to follow central exchange listing decisions. To do so, we construct a measure of a peripheral exchange's tendency to follow a central exchange:

Listing Following Probability_e =
$$\frac{\# \text{ Listings within } [0,30] \text{ days window}_e}{\# \text{ Listings}_e}$$
 (22)

Listing Following Probability_e simply takes the fraction of a peripheral exchange's coin listings which occured in a 30-day window after the central exchange listed the same coin. When Listing Following Probability_e is 1, in all cases where e lists a coin, the listing occurs from 0 to 30 days after the central exchange's listing. In implementing (22), we drop coins initially listed by central exchanges, and we drop peripheral exchanges with less than five listings in the sample period.

We then estimate a simple OLS specification:

$$ListingFollowingProbability_e = \alpha + \beta Correlation_e + \epsilon_e \tag{23}$$

If price correlations are associated with peripheral exchanges' listing following propensity, we expect β to be positive. We plot *Correlation*_e and *ListingFollowingProbability*_e and the slope of the estimated regression line, for Coinbase and Binance separately, in Figure 7. Both panels show a strong positive relationship, and the estimated β is positive and significant in both cases. Quantitatively, an exchange with an 0.01 higher BTC return correlation with a central exchange has a 0.11pp (0.09pp) higher probability of following the central exchange's listing decisions, respectively, for Binance (Coinbase).

6.6 Volume Decomposition

Our results imply that central exchange listings have three separate effects on total trade volumes: a direct effect from volume on the central exchange, and two indirect effects through increasing incumbents' volumes, and inducing new exchanges to list the coin. To estimate the relative sizes of these effects, we estimate three coin-level DID regressions:¹⁰

0.1

$$\log\left(TradeVol_{c,t}[inc]\right) = \sum_{k=-31}^{31} \left(\beta_k^{inc}\right) \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t} \quad (24)$$

$$\log\left(TradeVol_{c,t}[inc+ent]\right) = \sum_{k=-31}^{31} \left(\beta_k^{inc} + \beta_k^{ent}\right) \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t} \quad (25)$$

$$\log\left(TradeVol_{c,t}[inc+ent+cen]\right) = \sum_{k=-31}^{31} \left(\beta_k^{inc} + \beta_k^{ent} + \beta_k^{cen}\right) \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t} \quad (26)$$

The three dependent variables reflect increasingly broad definitions of total volume. In Specification (24), we use only volume on incumbent exchanges which listed the coin at least 30 days before the central exchange lists; in (25) we add entrant exchanges, which list after the central exchange; and in (26) we add the central exchange itself.

The top row of Figure 8 illustrates the estimated DID coefficients, which respectively reflect β_k^{inc} , $\beta_k^{inc} + \beta_k^{ent}$, and $\beta_k^{inc} + \beta_k^{ent} + \beta_k^{cen}$. As expected, the estimated coefficients are larger the broader the definition of volume that we use. Taking differences, we estimate β_k^{inc} , β_k^{ent} , and β_k^{cen} , which we plot in the second row; these have an intuitive interpretation as the percentage volume increases attributable to incumbents, entrants, and the central exchange respectively.

Quantitatively, we can calculate the percentage volume increase implied by the blue line in the top row of Figure 8, for any k by taking $\exp(\beta_k^{inc} + \beta_k^{cen} + \beta_k^{ent}) - 1$. Averaging over $0 \le k < 30$, we estimate that Binance listings are associated with an average 1606pp increase in total volume. The second row shows that this multiplicatively decomposes into a 94pp increase in incumbent volume, a 132pp increase associated with entrants, and a 292pp increase on Binance itself. Coinbase listings are associated with an average 331pp increase in total volumes within the first 30 days after listing. This multiplicatively decomposes into a 75pp increase in incumbent volumes increase, 78pp increase attributed to entrants, and 42pp on Coinbase itself.

We can then calculate a simple "external volume ratio":

$$EVR_k = \frac{\exp\left(\beta_k^{inc} + \beta_k^{ent}\right) - 1}{\exp\left(\beta_k^{cen}\right) - 1}$$
(27)

which measures how much "external" volume central exchange listings generate, on incumbent and entrant peripheral exchanges, for each dollar of volume directly generated on the central

¹⁰Here, we simply ignore observations with zero volumes, estimating only intensive margin effects; coin-day volumes are zero in around 15% of observations – this is slightly lower than the 20% of coin-exchange-day volumes that are missing, but the caveats in Chen and Roth (2024) apply.

exchange. We estimate the post-listing 30-day average of EVR_k to be 1.3 for Binance and 5.1 for Coinbase. EVR_k is smaller for Binance primarily because the estimated direct effect is larger; however, in both cases, our estimates of EVR_k imply that a large fraction of the effects of central exchanges listings on trade volumes arises through spillovers on peripheral exchanges.

7 Robustness Checks

7.1 Delistings

Our main analysis focuses only on new listings of coins. *Delistings* of coins from central exchanges are rarer; however, our model implies that delistings should have the opposite effect of new listings, decreasing volumes and listing propensity, and increasing price dispersion. We analyze delistings in Appendix H.1, verifying that these predictions mostly hold. The estimated effects are mostly statistically significant for Binance; for Coinbase, the estimated effects are sometimes insignificant, possibly because we only observe 17 delistings from Coinbase in our sample period.

7.2 Zero-Volume Observations

Trade and arbitrage volumes are frequently zero in our dataset; our baseline specifications handle this by separately estimating "intensive" and "extensive" margin effects for both variables, as suggested by Chen and Roth (2024). We show in Appendix H.2 that our results are also robust to other ways of handling zeros: in Appendix Tables A.4, A.5, and A.6, we repeat the specifications of Tables 3, 4, and 5 using the commonly used log (a + x) transformation for volume, for different choices of a; results mostly remain significant with unchanged sign, with the exception of some coefficient estimates in Table 5. Another approach is to subset the data to a balanced panel of coin-exchanges pairs which do not have zeros in dependent variables around listing events; we show estimates from this approach in Appendix Table A.7, finding that the arbitrage volume and trade volume results remain significant, though the correlation tests of Table 5 lose significance.

7.3 Other Economic Channels

Attention. An important alternative explanation for our results is that, instead of our arbitrage hypothesis, investors' *attention* to coins is a confounding variable which may influence

(or be influenced by) listings and trade volumes. We analyze the possible confounding effects of attention in Appendix H.3. Following studies such as Da, Engelberg and Gao (2011), we construct a coin-level measure of attention by collecting daily Google search volumes for the ticker symbols of coins in our sample.

We then analyze two ways in which investor attention could confound the interpretation of our results. First, we test what we call the "attention drives everything" hypothesis: variation in investor attention to coins is the core causal variable, driving central exchange listings, peripheral exchange listings, and trade volumes. In its strongest version, this hypothesis implies that central exchange listings have no causal effect on outcomes, in the sense that a central exchange exogeneously choosing to list a coin would not have any effect on outcomes: all changes are driven by gradual shifts in investors' attention driving all outcomes.

To test this, we build a matched sample in which treated and control coins have similar pre-listing attention characteristics. We find that coins listed on central exchanges experience increases in volume and listing propensity, and decreases in price dispersion, even relative to control coins with similar pre-listing volume level and attention trends, as shown in Appendix Table A.9 and A.10. This is difficult to reconcile with the strict "attention drives everything" hypothesis, in which central exchange listings have no causal effect on outcomes.

Second, attention may serve as a channel through which central exchange listings influence outcomes. We cannot fully rule out this channel – in fact, we find evidence suggesting it likely plays a role in the results. However, we provide evidence showing that it seems unlikely that the attention channel quantitatively explains the majority of the effect of listings on trade volumes.

Endorsement Effects. Another related channel is that central exchange listings may serve as "endorsements" of coins, enhancing their reputation. This omitted variable could contribute to explaining our trade volume increase and listing following results, though it is more difficult to explain how this channel could explain our arbitrage flow and price dispersion results. In principle, large-cap coins with greater existing market adoption should be less affected by the endorsement channel. Thus, we conduct a robustness check in which we restrict our sample to treated coins with a market capitalization of more than \$100 million one month prior to their central exchange listing date; this approximately corresponds to selecting the top tertile of treated coins. Results are shown in Appendix Tables A.15 and A.16; the estimated effects are significant, and quantitatively similar to our baseline estimates.

Illegal Activities. Another concern is that some exchanges may primarily serve as marketplaces for money laundering, drug trade, and other illegal activities. To check whether our trade volume results are driven mainly by these exchanges, we follow Amiram, Lyandres

and Rabetti (2022) and collect data on exchanges' Anti-Money Laundering/Know Your Customer (AML/KYC) requirements from coinintelligence.com. We then estimate our trade volume specifications on a subsample of 59 exchanges that impose both AML and KYC requirements. We show results in Appendix Table A.17: results remain statistically significant, with unchanged signs and similar magnitudes to our baseline specifications.

Fake Volume. Trade volumes reported by exchanges are known to be subject to manipulation and wash trading (Cong et al., 2023b). One particular concern is if, in response to a central exchange's entry decision, smaller exchanges increasingly engage in wash trading, as suggested by Amiram, Lyandres and Rabetti (2022). While this hypothesis can partly account for our trading volume results, it is unlikely to affect the on-chain arbitrage volume results, which are costly to manipulate, as well as our other results.

To provide further evidence that volume falsification is not the main driver of our results, we restrict to 29 exchanges analyzed in Cong et al. (2023b) and analyze the responses of volumes in their three groups of exchanges.¹¹ The classification reflects the intensity, to what extent the exchanges are regulated and thus falsely report their trading volume. The regulated exchanges are supervised by New York State Department of Financial Services (NYSDFS), and they are classified into regulated exchanges. The unregulated exchanges are further classified into two different tiers based on their web traffic. Cong et al. (2023b) argue that wash trading should be most prevalent in tier-2 exchanges, less prevalent in tier-1 exchanges, and least prevalent in regulated exchanges.

The intensive margin effects are shown in Appendix Table A.18. The estimates are statistically significant, and similar in magnitude, for both tier-2 exchanges, for which "fake volume" issues are comparatively severe, and tier-1 exchanges, for which they are less severe. The intensive margin effects for regulated exchanges are statistically insignificant; however, this group consists of only two exchanges besides Coinbase – Bitstamp and Gemini – hence our standard errors are quite large for this group. The extensive margin effects are shown in Appendix Table A.19. We find positive and significant effects on the extensive margin for tier-1 exchanges, but insignificant effects for regulated and tier-2 exchanges. Note that the extensive margin is much less relevant for regulated exchanges, since only 4% of regulated coin-exchange-days have zero volumes, relative to 20% in the full sample.

¹¹These exchanges are: (1) regulated exchanges: Bitstamp, Coinbase, Gemini; (2) tier-1 unregulated exchanges: Binance, Bittrex, Bitfinex, Hitbtc, Huobi, Kucoin, Liquid, Okex, Poloniex, Zb; (3) tier-2 unregulated exchanges: Bgogo, Biki, Bitz, Coinbene, Dragonex, Lbank, Mxc, Fcoin, Exmo, Coinmex, Bibox, Bitmart, Bitmax, Coinegg, Digifinex, and Gateio.

7.4 Subsample Analysis

Another concern is whether our findings are specific to a particular time period. We separately estimate our effects for two subsamples: (1) January 2017 to July 2022, representing an early boom-bust cycle, and (2) August 2022 to December 2023, a period of market growth and increasing crypto institutionalization. The results for price dispersion, arbitrage volumes, and listing following tendency are presented in Appendix Tables A.20 and A.21, and the trade volume results are shown in Appendix Tables A.22 and A.23. Results are qualitatively and quantitatively similar, though the estimated effects on price dispersion are insignificant in the latter part of our sample, potentially because we have lower power in the latter part of the sample.

8 Conclusion

The global market for cryptocurrency exchanges is fragmented. This fragmentation does not appear to be a simple story of competition over customer trade volumes. Instead, our results suggest that small peripheral exchanges play a role akin to brokers or dealers in traditional financial markets, giving their customers access to global crypto market liquidity through their connections to a few large exchanges, which play a role similar to inter-dealer markets for traditional assets.

This market structure implies that the listing decisions of large exchanges play an outsized role in shaping trade volumes and prices in the global cryptocurrency market. Exchange listings are a point of potential regulatory oversight: large exchange listings decisions are unregulated or lightly regulated in many jurisdictions, and it is not clear that large exchanges' listing incentives are aligned with the interests of market participants.

Our results imply that *arbitrageurs* are a second potential target for regulation. Market participants playing similar roles in traditional financial markets are subject to strict reporting and compliance requirements: for instance, authorized participants in ETF markets must comply with reporting requirements,¹² and designated market makers (DMMs) face performance standards and trading restrictions.¹³ To better monitor market conditions in crypto markets, regulators could require crypto exchanges to report on their arbitrageurs' trading activity and sources of liquidity. Arbitrageur regulation is also a potential policy lever for modulating crypto market efficiency: regulators who wish to lower frictions in local crypto markets could decrease arbitrageurs' barriers to entry, whereas regulators who wish

¹²See 17 CFR § 270.6c-11

¹³Rules and restrictions applied to DMMs can be found in the NYSE Rules; see for example Rule 98.

to curtail access to global crypto markets could do so by imposing costs and restrictions on arbitrageurs in addition to exchanges.

To conclude, we tentatively ask: why is the "networks of linked exchanges" market structure we analyze here so rare in markets for traditional financial assets?

A narrative answer to this question, extrapolating from our results, is that this difference in market structures is driven by primitive technical differences between cryptocurrencies and traditional financial assets. Traditional assets are legal constructs, which mostly exist within singular jurisdictional and fiat ecosystems. The role of a US equity exchange is to match buy and sell orders from customers with some form of US presence; the equity exchange can largely assume cash settlement will occur through the US banking system and equity settlement through the DTCC. Traditional financial markets are not always fully centralized: there are many competing equity exchanges in the US; markets for bonds and derivatives have a core-periphery OTC structure. But the relative unification of infrastructure may be a reason why customer bases are not naturally segmented enough for a "network of local exchanges" market structure to emerge.¹⁴

Cryptocurrencies are not legal constructs: they are entries in consensus-based ledgers, which can be "held" and "sent" wherever users can access the internet, and fungible regardless of where in the world a wallet owner resides. But customers in different countries use different fiat payment systems, and face different regulatory constraints on financial market participation. As in commodity markets, the anachronistic "network of exchanges" market structure emerges naturally in response to these primitives: local crypto exchanges match orders within jurisdictions, and the "transportation costs" of bridging customer groups are borne by specialized arbitrageurs.

Even more tentatively, what implications do our results have for the future evolution of cryptocurrency exchange market structure? Our narrative suggests that full cryptocurrency exchange consolidation in the near future is unlikely: some degree of fragmentation is likely to persist as long as regulatory and fiat segmentation across customers persists. To the extent that fiat rails are *replaced* with cryptocurrency or stablecoin-based systems in the longer run, though, some of the boundaries which separate customer groups at present may loosen in some parts of the world, leading to potentially greater consolidation of cryptocurrency

¹⁴One possible exception is a small number of dual-listed companies, which trade on more than one exchange at once, analyzed in Rosenthal and Young (1990), Froot and Dabora (1999), De Jong, Rosenthal and Van Dijk (2009), and Vandeweyer, Yang and Yannelis (2024). However, dual-listed companies are rare, and trade on a small number of exchanges relative to cryptocurrencies, making it essentially impossible for core-periphery network structure to emerge. Another possible exception is ADRs, analyzed in Gagnon and Karolyi (2010); however, ADRs are not fungible with their underlying stocks in the same way cryptocurrencies are across exchanges, and ADRs also trade on relatively few exchanges at once.

exchanges in the long run.

References

- Admati, Anat R, and Paul Pfleiderer. 1988. "A theory of intraday patterns: Volume and price variability." *The Review of Financial Studies*, 1(1): 3–40.
- Amiram, Dan, Evgeny Lyandres, and Daniel Rabetti. 2022. "Cooking the order books: information manipulation and competition among crypto exchanges." Available at SSRN 3745617.
- Ante, Lennart. 2019. "Market reaction to exchange listings of cryptocurrencies." Blockchain Research Lab Working Paper Series.
- Augustin, Patrick, Alexey Rubtsov, and Donghwa Shin. 2023. "The impact of derivatives on spot markets: Evidence from the introduction of bitcoin futures contracts." *Management Science*, 69(11): 6752–6776.
- Babus, Ana, and Cecilia Parlatore. 2022. "Strategic fragmented markets." Journal of Financial Economics, 145(3): 876–908.
- Babus, Ana, and Péter Kondor. 2018. "Trading and information diffusion in over-thecounter markets." *Econometrica*, 86(5): 1727–1769.
- Babus, Ana, and Tai-Wei Hu. 2017. "Endogenous intermediation in over-the-counter markets." *Journal of Financial Economics*, 125(1): 200–215.
- Baldauf, Markus, and Joshua Mollner. 2021. "Trading in fragmented markets." Journal of Financial and Quantitative Analysis, 56(1): 93–121.
- Benedetti, Hugo, and Ehsan Nikbakht. 2021. "Returns and network growth of digital tokens after cross-listings." *Journal of Corporate Finance*, 66: 101853.
- Biais, Bruno, Thierry Foucault, and Sophie Moinas. 2015. "Equilibrium fast trading." Journal of Financial Economics, 116(2): 292–313.
- Borri, Nicola, and Kirill Shakhnov. 2022. "The cross-section of cryptocurrency returns." The Review of Asset Pricing Studies, 12(3): 667–705.
- Borri, Nicola, and Kirill Shakhnov. 2023. "Cryptomarket discounts." Journal of International Money and Finance, 139: 102963.
- Budish, Eric, Peter Cramton, and John Shim. 2015. "The high-frequency trading arms race: Frequent batch auctions as a market design response." The Quarterly Journal of Economics, 130(4): 1547–1621.
- Budish, Eric, Robin S Lee, and John J Shim. 2024. "A theory of stock exchange competition and innovation: Will the market fix the market?" *Journal of Political Economy*, 132(4): 1209–1246.
- Chang, Briana, and Shengxing Zhang. 2015. "Endogenous market making and network formation."
- Chao, Yong, Chen Yao, and Mao Ye. 2019. "Why discrete price fragments US stock exchanges and disperses their fee structures." *The Review of Financial Studies*, 32(3): 1068–1101.
- Chen, Daniel, and Darrell Duffie. 2021. "Market fragmentation." American Economic Review, 111(7): 2247–74.
- Chen, Jiafeng, and Jonathan Roth. 2024. "Logs with zeros? Some problems and solutions." The Quarterly Journal of Economics, 139(2): 891–936.
- Choi, Kyoung Jin, Alfred Lehar, and Ryan Stauffer. 2022. "Bitcoin microstructure and the kimchi premium." Available at SSRN 3189051.
- Colliard, Jean-Edouard, and Thierry Foucault. 2012. "Trading fees and efficiency in limit order markets." The Review of Financial Studies, 25(11): 3389–3421.
- Cong, Lin William, Ke Tang, Yanxin Wang, and Xi Zhao. 2023a. "Inclusion and democratization through web3 and defi? Initial evidence from the ethereum ecosystem." National Bureau of Economic Research Cambridge W30949. Series: Working Paper Series.
- Cong, Lin William, Xi Li, Ke Tang, and Yang Yang. 2023b. "Crypto wash trading." Management Science, 69(11): 6427–6454.
- Craig, Ben, and Yiming Ma. 2022. "Intermediation in the interbank lending market." Journal of Financial Economics, 145(2): 179–207.
- Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. "In search of attention." The Journal of Finance, 66(5): 1461–1499.
- **De Jong, Abe, Leonard Rosenthal, and Mathijs A Van Dijk.** 2009. "The risk and return of arbitrage in dual-listed companies." *Review of Finance*, 13(3): 495–520.
- Di Maggio, Marco, Amir Kermani, and Zhaogang Song. 2017. "The value of trading relations in turbulent times." *Journal of Financial Economics*, 124(2): 266–284.

- **Dugast, Jérôme, Semih Úslü, and Pierre-Olivier Weill.** 2022. "A theory of participation in OTC and centralized markets." *The Review of Economic Studies*, 89(6): 3223–3266.
- **Dunne, Peter G, Harald Hau, and Michael J Moore.** 2015. "Dealer intermediation between markets." *Journal of the European Economic Association*, 13(5): 770–804.
- **Dyhrberg, Anne Haubo.** 2020. "One coin, many markets-how market frictions affect arbitrageurs." Working paper, https://www.annedyhrberg.org/docs/Dyhrberg-2020.pdf.
- Farboodi, Maryam. 2023. "Intermediation and voluntary exposure to counterparty risk." Journal of Political Economy, 131(12): 3267–3309.
- Farboodi, Maryam, Gregor Jarosch, and Robert Shimer. 2023. "The emergence of market structure." *The Review of Economic Studies*, 90(1): 261–292.
- Foucault, Thierry, and Albert J Menkveld. 2008. "Competition for order flow and smart order routing systems." *The Journal of Finance*, 63(1): 119–158.
- Froot, Kenneth A, and Emil M Dabora. 1999. "How are stock prices affected by the location of trade?" *Journal of Financial Economics*, 53(2): 189–216.
- Gagnon, Louis, and G Andrew Karolyi. 2010. "Multi-market trading and arbitrage." Journal of Financial Economics, 97(1): 53–80.
- Hagströmer, Björn, and Albert J Menkveld. 2019. "Information revelation in decentralized markets." *The Journal of Finance*, 74(6): 2751–2787.
- Hautsch, Nikolaus, Christoph Scheuch, and Stefan Voigt. 2024. "Building trust takes time: limits to arbitrage for blockchain-based assets." *Review of Finance*, 28(4): 1345–1381.
- Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff. 2020. "Relationship trading in over-the-counter markets." *The Journal of Finance*, 75(2): 683–734.
- Hoopes, Jeffrey L, Daniel H Reck, and Joel Slemrod. 2015. "Taxpayer search for information: Implications for rational attention." *American Economic Journal: Economic Policy*, 7(3): 177–208.
- Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill. 2022. "Heterogeneity in decentralized asset markets." *Theoretical Economics*, 17(3): 1313–1356.
- Klemperer, Paul D, and Margaret A Meyer. 1989. "Supply function equilibria in oligopoly under uncertainty." *Econometrica: Journal of the Econometric Society*, 57(6): 1243– 1277.

- Kroeger, Alexander, and Asani Sarkar. 2017. "The law of one bitcoin price." Federal Reserve Bank of Philadelphia.
- Li, Dan, and Norman Schürhoff. 2019. "Dealer networks." The Journal of Finance, 74(1): 91–144.
- Li, Jiasun, Mei Luo, Muzhi Wang, and Zhe Wei. 2022. "Cryptocurrency Listings on Cryptocurrency Exchanges." Donald G. Costello College of Business at George Mason University Research Paper.
- Li, Jun, Xianwei Liu, Qiang Ye, Feng Zhao, and Xiaofei Zhao. 2023. "It depends on when you search." *MIS Quarterly*, 47(1): 263–280.
- Makarov, Igor, and Antoinette Schoar. 2019. "Price discovery in cryptocurrency markets." AEA Papers and Proceedings, 109: 97–99.
- Makarov, Igor, and Antoinette Schoar. 2020. "Trading and arbitrage in cryptocurrency markets." *Journal of Financial Economics*, 135(2): 293–319.
- Makarov, Igor, and Antoinette Schoar. 2021. "Blockchain analysis of the bitcoin market." Available at SSRN 3942181.
- Malamud, Semyon, and Marzena Rostek. 2017. "Decentralized exchange." American Economic Review, 107(11): 3320–62.
- Newman, Mark EJ. 2008. "The mathematics of networks." The new palgrave encyclopedia of economics, 2(2008): 1–12.
- O'Hara, Maureen, and Mao Ye. 2011. "Is market fragmentation harming market quality?" *Journal of Financial Economics*, 100(3): 459–474.
- Pagano, Marco. 1989. "Trading volume and asset liquidity." The Quarterly Journal of Economics, 104(2): 255–274.
- **Pesaran, M Hashem.** 2006. "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, 74(4): 967–1012.
- Rosenthal, Leonard, and Colin Young. 1990. "The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever NV/PLC." Journal of Financial Economics, 26(1): 123–141.
- Rostek, Marzena, and Ji Hee Yoon. 2021. "Exchange design and efficiency." *Econometrica*, 89(6): 2887–2928.

- Rostek, Marzena J, and Ji Hee Yoon. 2020. "Equilibrium theory of financial markets: Recent developments." Available at SSRN 3710206.
- Shams, Amin. 2020. "The structure of cryptocurrency returns." Fisher College of business working paper.
- Talamas, Roberto. 2021. "Analyzing the Crypto Exchange Pump Phenomenon." Messari.
- Tsang, Kwok Ping, and Zichao Yang. 2020. "Price dispersion in bitcoin exchanges." *Economics Letters*, 194: 109379.
- **Üslü, Semih.** 2019. "Pricing and liquidity in decentralized asset markets." *Econometrica*, 87(6): 2079–2140.
- Vandeweyer, Quentin, Minghao Yang, and Constantine Yannelis. 2024. "Discount factors and monetary policy: Evidence from dual-listed stocks." National Bureau of Economic Research W32499. Series: Working Paper Series.
- van Kampen, Reinout. 2022. "Cross-listings: Cryptocurrency Characteristics and Abnormal Returns." Master Thesis.
- Viswanathan, S, and James JD Wang. 2004. "Inter-dealer trading in financial markets." The Journal of Business, 77(4): 987–1040.
- Wang, Chaojun. 2016. "Core-periphery trading networks." Available at SSRN 2747117.
- Wittwer, Milena. 2021. "Connecting disconnected financial markets?" American Economic Journal: Microeconomics, 13(1): 252–282.
- Wu, Xian. 2024. "Dynamic Market Choice." Working Paper.
- Yoon, Ji Hee. 2018. "Endogenous market structure: Over-the-counter versus exchange trading." Working paper.
- Yu, Gloria Yang, and Jinyuan Zhang. 2022. "Flight to bitcoin." Available at SSRN 3278469.

Figure 1: Equilibrium Outcomes in Example Networks

Equilibrium prices and arbitrage flows for two example networks in our model. Panel A shows two exchanges connecting by a single arbitrageur. Panel B shows four "peripheral" exchanges 1, 2, 3, 4 arranged around a central exchange C. In both cases, we set $\gamma = 2, \zeta = 2, \psi = 0$. Analytical solutions are derived in Appendices B.1 and B.2 respectively.

(a) Two Nodes

$$\begin{array}{c}
x_1 = 2 \\
x_1 - z_{12} = 1 \\
p_1 = -2
\end{array}$$
(b) Multiple Nodes
(c) Multiple Nodes

$$\begin{array}{c}
x_2 = 1 \\
x_2 - z_{2C} = 0.583 \\
p_2 = -1.167
\end{array}$$

$$\begin{array}{c}
x_2 = 1 \\
x_2 - z_{2C} = 0.583 \\
p_2 = -1.167
\end{array}$$

$$\begin{array}{c}
x_C = 0 \\
x_C + \sum_{i=1}^{4} z_{iC} = 0.167 \\
p_C = -0.333
\end{array}$$

$$\begin{array}{c}
x_4 = -2 \\
x_4 - z_{4C} = -0.917 \\
p_4 = 1.833
\end{array}$$

$$\begin{array}{c}
x_1 = 3 \\
x_2 = 1 \\
x_2 - z_{2C} = 0.583 \\
z_{3C} = -0.583
\end{array}$$

$$\begin{array}{c}
x_3 = -1 \\
x_3 - z_{3C} = -0.417 \\
p_3 = 0.833
\end{array}$$

Figure 2: Equilibrium Connection Strategies

This figure shows the actions available to an arbitrageur in our model, assuming other arbitrageurs coordinate on a star-network equilibrium. The black lines denote links formed by the other arbitrageurs, connecting all but one of the peripheral exchanges to a single central exchange: the gray dot represents the isolated exchange. The red, blue, orange, and green strategies denote strategies that the last remaining arbitrageur can take: she can compete with an existing arbitrageur over a peripheral exchange (red), link the isolated exchange to a peripheral exchange (green), link two connected peripheral exchanges (orange), or complete the star network by linking the isolated exchange to the central exchange (blue). A star-network equilibrium exists for any parameter settings such that, conditional on all other arbitrageurs forming a star network, the final arbitrageur finds it optimal to complete the star, so the blue strategy has higher expected profit than the other three strategies. Letters correspond to the node labels used in the proof of Proposition 1 in Appendix C.



This figure shows the distribution of the pairwise correlations of BTC returns, for all exchange pairs excluding pairs involving Binance or Coinbase, and exchange pairs between peripheral exchanges and either Binance or Coinbase. For each exchange pair, we calculate return correlations using the entire time period where we have coverage for both exchanges in the pair. Data source: Cryptotick.



Figure 4: Price Gap Reversion

This figure shows, from the top to bottom row, the distribution of estimates from Specifications (14), (15), and (16) for all coin-exchange pairs:

$$\Delta PriceGap_{c,e,t} = \beta_{c,e}^{PriceGap} PriceGap_{c,e,t-1} + \epsilon_{c,e,t}^{PriceGap}$$
$$\Delta p_{c,e,t} = \beta_{c,e}^{p} PriceGap_{c,e,t-1} + \epsilon_{c,e,t}^{p}$$
$$\Delta p_{c,t}^{cen} = \beta_{c,e}^{p^{cen}} PriceGap_{c,e,t-1} + \epsilon_{c,e,t}^{p^{cen}}$$

In all cases, one data point is one peripheral exchange-coin pair. The top row shows the distribution of estimated $\beta_{c,e}^{PriceGap}$, where the dependent variable is the change in price gap of coin c between peripheral exchange e and central exchange (either Binance or Coinbase) at day t. The red vertical line shows $\beta^{PriceGap} = -1$, the zero-persistence benchmark, where price gaps on a given day fully revert on the following day. The middle row shows the distribution of estimated $\beta_{c,e}^{p}$, where the dependent variable is the price change in coin c on peripheral exchange e at day t. The bottom row shows the distribution of estimated $\beta_{c,e}^{p^{cen}}$, and the dependent variable is the price change in coin c on central exchange (either Binance or Coinbase) at day t. The red vertical lines in the middle and bottom row denote $\beta_{c,e}^{p} = 0$ and $\beta_{c,e}^{p^{cen}} = 0$, respectively. Data source: Cryptotick.



Figure 5: Central Exchange Listings and Price Dispersion

This figure shows how the interquartile range of coin prices changes around listing events. For each coin affected by a listing, we measure the median, 25th, and 75th percentile prices across incumbent exchanges, and normalize all percentiles by the median. Incumbent exchanges are defined as exchanges that list coin c at least 30 days before its listing on a central exchange. The blue, green, and red lines represent the average normalized prices at the 75th, median, and 25th percentiles across coins, respectively. The x-axis represents the time interval from the central exchange's listing date, with the red vertical line indicating the listing time. For central exchanges, we focus on Binance and Coinbase. Data source: Cryptotick.



Figure 6: Listing Times of Peripheral Exchanges Relative to Central Exchanges

This figure shows the histogram of peripheral exchange listing times relative to central exchange listings, including only coins that were listed on any exchange for at least 30 days before being listed on central exchanges. The x-axis represents the time interval from the central exchange's listing date, with the red vertical line indicating the listing time. The y-axis shows the mass of each time interval bar. For central exchanges, we focus on Binance and Coinbase. Data source: Cryptotick.



Figure 7: Price Correlations and Listing Following Tendency

This figure shows the relationship between BTC return correlation and listing following probability for all exchanges relative to central exchanges. Each data point represents an exchange. The x-axis denotes the return correlation of Bitcoin between an exchange and a central exchange using the entire time period where we have coverage for both exchanges in the pair. The y-axis denotes the listing following probability between an exchange and a central exchange, defined in Equation (22). The red line indicates the fitted linear regression curve. The correlation coefficient and its p-value are reported. For central exchanges, we focus on Binance and Coinbase. Data source: Cryptotick.



Figure 8: Volume Increase Decomposition

The top row of the figure shows estimates from Specification (24) to (26):

$$\log(Volume_{c,t}[inc]) = \sum_{k=-31}^{31} \left(\beta_k^{inc}\right) \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$
$$\log(Volume_{c,t}[inc+ent]) = \sum_{k=-31}^{31} \left(\beta_k^{inc} + \beta_k^{ent}\right) \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$
$$\log(Volume_{c,t}[inc+ent+cen]) = \sum_{k=-31}^{31} \left(\beta_k^{inc} + \beta_k^{ent} + \beta_k^{cen}\right) \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$

along with 95% confidence intervals. We categorize exchange-coin pairs into three groups according to their listing time and central status: incumbent (exchanges that list at least 30 days before central exchange listings or coins that are not listed by central exchanges), central (central exchange), and entrant (all remaining exchanges). The dependent variable is coin-day level log dollarized trade volume, computed by filtering the relevant exchange-coin pairs, adding up the dollarized volume, and taking the log. In the top row, the red line shows estimates of β_k^{inc} , using only incumbent volume as the dependent variable. The green line shows estimates of $(\beta_k^{inc} + \beta_k^{ent})$, using incumbent plus entrant volume as the dependent variable. The blue line shows estimates of $(\beta_k^{inc} + \beta_k^{ent} + \beta_k^{cen})$, using incumbent, entrant, and central exchange volume as the dependent variable. The blue line shows estimates coin fixed effects, and η_t represents day fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin and time level. We then recover $\beta_k^{inc}, \beta_k^{ent}$, and β_k^{cen} simply by taking differences between the estimated coefficients. In the bottom row, the red line shows estimates of β_k^{ent} , Data source: Cryptotick.



Table 1: Summary Statistics

This table shows summary statistics for variables related to our Cryptotick and Ethereum on-chain flow data. Panel A presents descriptive statistics for the exchange level BTC return correlation and the listing-following probability with respect to Binance and Coinbase. Panel B summarizes the coin-day level variables, Panel C presents the coin-exchange-day level variables, and Panel D displays the arbitrage flow amounts at different levels. For each variable, we show the number of non-missing observations, the mean, the standard deviation and the 25th, 50th and 75th percentile values. Data source: Cryptotick, Etherscan.

Panel A: Exchange L	evel											
		with Binance						with Coinbase				
	Obs.	Mean	n SD	p25	p50	p75	Obs.	Mean	SD	p25	p50	p75
BTC Return Correlation Listing Following Prob	$\begin{array}{c} 241 \\ 164 \end{array}$	$0.81 \\ 0.09$	$0.2 \\ 0.1$	$\begin{array}{c} 0.76 \\ 0 \end{array}$	$0.86 \\ 0.059$	$\begin{array}{c} 0.94 \\ 0.15 \end{array}$	242 158	$0.76 \\ 0.075$	$\begin{array}{c} 0.24\\ 0.081 \end{array}$	$\begin{array}{c} 0.65\\ 0 \end{array}$	0.83 0.061	0.94 0.12
Panel B: Coin-day	Level											
	Obs.		Mea	n	S	D	p25		p50)	p'	75
Price Dispersion	868132	8132		0.12		.2	0.011		0.03		0.14	
I(ArbVol > 0)	272574	2574		0.63		48	0		1		1	
Log(ArbVol)	171601	71601		11		3	8.8		11		13	
Net Listings	1353455		0.0048		0.	16	0		0		0	
Panel C: Coin-exch	ange-day i	Level										
	Obs.		Μ	ean		SD	p25	5	p5	50	I	o75
I(TradeVol > 0)	12034327		0.81			0.39	1		1			1
Log(TradeVol)	9747409		11			4.2	8.9	12			14	
Panel D: Arbitrage	Transacti	ons										
				Obs.	М	lean	SD	p25	5	p50	p	75
ArbVol (Transaction Level, \$)			21	222854	40	0078	539773	485	5	2002	101	64
ArbVol (Coin Level, \$Millions)				239	3	559	25755	5.7	7	80	49	93
ArbVol (Exchange Leve	l, \$Millions)		93	9	146	30428	11		261	17	19
ArbVol (Exchange-pair Level, \$Millions)				2360	1	.80	1820	0.00	83	0.12	2	.1

Table 2: Core-Periphery Structure of Arbitrage Flows

This table shows the arbitrage volume market share, trade volume market share, degree centrality, betweenness centrality, and eigenvector centrality, for the 10 exchanges with the highest arbitrage volume. Exchanges' market share is calculated from their arbitrage volume $ArbVol_e$ and trade volume $TradeVol_e$. $TradeVol_e$ is the total trade volume for all coins, fiat, and stablecoins in our Cryptotick data. We assume two exchanges e and \tilde{e} are "connected" whenever the arbitrage flows between e and \tilde{e} are a large fraction of total arbitrage flows involving either exchange: formally, if $ArbVol_{e,\tilde{e}}$, is at least 10% of the smaller of $ArbVol_e$ and $ArbVol_{\tilde{e}}$. Degree centrality for each exchange is defined as the number of edges it has with other exchanges. Betweenness centrality counts the fraction of shortest paths between two randomly chosen exchanges that pass through a given exchange. Eigenvector centrality captures global importance in contrast to degree centrality. All three centrality measures are normalized so that the total centrality across all exchanges sums to one. Data source: Cryptotick, Etherscan.

Exchange	ArbVol Share	TradeVol Share	Degree	Betweenness	Eigenvector
Binance	0.25	0.21	0.2	0.68	0.08
FTX	0.15	0.01	0.04	0.03	0.02
Coinbase	0.13	0.03	0.05	0.06	0.03
Huobi	0.1	0.05	0.06	0.06	0.03
Kraken	0.1	0.01	0.04	0.06	0.02
Bitfinex	0.07	0.01	0.04	0.04	0.02
OKEX	0.06	0.05	0.04	0.02	0.02
Kucoin	0.02	0.01	0.01	0	0.01
BinanceUS	0.01	0.01	0.01	0	0.01
Bithumb	0.01	0.02	0.01	0	0.01

Table 3: Central Exchange Listings: Coin Level

This table shows estimates from Specification (17):

$$\begin{aligned} Y_{c,t} = & \beta_1 Listing(0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayListing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{aligned}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For price dispersion and arbitrage volume, we exclude newly entered exchanges. For coin listings, we include only coins that were listed on any exchange for at least 30 days before being listed on central exchanges or were never listed by central exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan.

Dependent Variables:	Dispersion		I(ArbV	Vol > 0)	Log(A	rbVol)	Net Listings		
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Listing (0-30 days)	-0.04***	-0.04***	0.41***	0.25***	2.2^{***}	2.0^{***}	0.08***	0.08***	
	(0.01)	(0.008)	(0.05)	(0.03)	(0.28)	(0.23)	(0.01)	(0.01)	
Listing $(> 30 \text{ days})$	-0.09***	-0.06***	0.42^{***}	0.25^{***}	1.9^{***}	1.7^{***}	0.005^{*}	-0.009***	
	(0.01)	(0.01)	(0.05)	(0.04)	(0.23)	(0.28)	(0.002)	(0.003)	
Pre Three-day Listing	-0.02	-0.02**	0.15^{**}	0.19^{***}	1.1^{***}	1.4^{***}	0.02^{*}	0.04^{*}	
	(0.01)	(0.009)	(0.06)	(0.04)	(0.29)	(0.25)	(0.01)	(0.02)	
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Adjusted R ²	0.50	0.47	0.42	0.40	0.54	0.52	0.06	0.07	
Observations	$645,\!346$	$847,\!421$	$200,\!044$	$256,\!331$	116,734	$156{,}578$	$1,\!224,\!478$	$1,\!341,\!152$	

This table shows estimates from Specification (18):

$$Y_{c,e,t} = \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$

The dependent variables are trade volume dummy and log trade volume. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	I(TradeVol > 0)				Log(TradeVol)				
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Listing (0-30 days)	0.02**	0.02***	0.02***	0.02***	0.74***	0.78***	0.81***	0.77***	
	(0.008)	(0.006)	(0.006)	(0.004)	(0.07)	(0.05)	(0.06)	(0.04)	
Listing $(> 30 \text{ days})$	0.06^{***}	0.04^{***}	0.05^{***}	0.03^{***}	0.72^{***}	0.86^{***}	0.82^{***}	0.85^{***}	
	(0.01)	(0.006)	(0.007)	(0.004)	(0.10)	(0.06)	(0.08)	(0.05)	
Pre Three-day Listing	0.01	0.03***	0.01^{*}	0.02***	0.62^{***}	1.0^{***}	0.69^{***}	1.0^{***}	
	(0.009)	(0.006)	(0.006)	(0.004)	(0.09)	(0.08)	(0.08)	(0.07)	
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Exchange FE \times Day FE	No	No	Yes	Yes	No	No	Yes	Yes	
Adjusted \mathbb{R}^2	0.47	0.48	0.64	0.70	0.79	0.78	0.84	0.84	
Observations	$4,\!895,\!893$	$9,004,\!676$	$4,\!895,\!893$	$9,\!004,\!676$	$3,\!790,\!310$	$7,\!220,\!758$	$3,\!790,\!310$	$7,\!220,\!758$	

Table 5: Trade Volume, Exchange Correlations, and Listings

This table shows estimates of Specification (21):

$$\begin{split} Y_{c,e,t} = & \beta_1 Listing (0\text{--}30 \text{ days})_{c,t} + \beta_2 Listing (0\text{--}30 \text{ days})_{c,t} \times Correlation_e + \\ & \beta_3 Listing (> 30 \text{ days})_{c,t} + \beta_4 Listing (> 30 \text{ days})_{c,t} \times Correlation_e + \\ & \beta_5 PreThreeday Listing_{c,t} + \beta_6 PreThreeday Listing_{c,t} \times Correlation_e + \\ & \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variables are trade volume dummy and log trade volume. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. $Correlation_e$ is the return correlation of Bitcoin between the central exchange and the peripheral exchange e using the entire time period where we have coverage for both exchanges in the pair. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:		I(TradeVol > 0)				Log(TradeVol)				
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase		
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
Listing (0-30 days)	-0.01	-0.06	0.10	0.09***	-3.2***	-1.2***	-0.93	0.39		
	(0.08)	(0.04)	(0.06)	(0.03)	(0.90)	(0.38)	(0.80)	(0.34)		
Listing $(0-30 \text{ days}) \times \text{Correlation}$	0.04	0.11^{**}	-0.09	-0.09**	4.3^{***}	2.3^{***}	1.9^{**}	0.43		
	(0.09)	(0.05)	(0.07)	(0.04)	(0.97)	(0.43)	(0.87)	(0.39)		
Listing $(> 30 \text{ days})$	-0.13	-0.37***	0.21^{***}	0.12^{***}	-6.4***	-2.3***	-2.3**	1.3^{***}		
	(0.10)	(0.05)	(0.08)	(0.04)	(1.1)	(0.49)	(0.91)	(0.45)		
Listing $(> 30 \text{ days}) \times \text{Correlation}$	0.21^{*}	0.49^{***}	-0.17^{**}	-0.11**	7.8***	3.7^{***}	3.4^{***}	-0.52		
	(0.11)	(0.06)	(0.08)	(0.04)	(1.2)	(0.55)	(1.0)	(0.50)		
Pre Three-day Listing	-0.06	-0.007	0.009	0.13^{***}	-3.6***	-0.86**	-1.1	0.67^{*}		
	(0.10)	(0.05)	(0.08)	(0.04)	(0.99)	(0.40)	(0.85)	(0.37)		
Pre Three-day Listing \times Correlation	0.08	0.05	0.003	-0.13***	4.6^{***}	2.2^{***}	1.9^{**}	0.39		
	(0.10)	(0.05)	(0.08)	(0.04)	(1.1)	(0.46)	(0.92)	(0.42)		
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		
Exchange FE \times Day FE	No	No	Yes	Yes	No	No	Yes	Yes		
Adjusted R ²	0.47	0.48	0.64	0.70	0.79	0.78	0.84	0.84		
Observations	$4,\!864,\!077$	$8,\!975,\!799$	$4,\!864,\!077$	$8,\!975,\!799$	3,771,526	$7,\!211,\!809$	3,771,526	$7,\!211,\!809$		

Internet Appendix

A Logistics of Transferring Cryptocurrencies Across Exchanges

In this appendix, we describe the process of arbitrage across crypto exchanges, illustrating some of the costs, delays, and risks involved. This builds on similar discussions in papers such as Makarov and Schoar (2020), Dyhrberg (2020), Choi, Lehar and Stauffer (2022), Augustin, Rubtsov and Shin (2023), and Hautsch, Scheuch and Voigt (2024).

Suppose a price gap emerges between ETH on a U.S. exchange (denominated in USD) and a Korean exchange (denominated in KRW), with the U.S. price being lower. Exploiting this arbitrage requires accounts on both exchanges, access to U.S. and Korean fiat, and the ability to exchange USD for KRW. The arbitrage process follows these steps:

- 1. Fiat USD is deposited at the US exchange, using a bank transfer or other means. The fiat is used to purchase ETH at the lower US price.
- 2. The purchased ETH is withdrawn from the US exchange and deposited in the Korean exchange.
- 3. The ETH is sold in Korean won at the higher Korean exchange price.
- 4. To convert the profits back to fiat USD, the Korean won is then withdrawn from the Korean exchange and exchanged for USD.

Several fees, restrictions, and delays affect the viability of this trade.

Fees. The total fees charged in the course of this transaction include fees charged by exchanges for depositing, trading, and withdrawing in steps 1, 2, and 3, as well as transaction fees charged for the blockchain transfer in step 2. Exchange fees vary widely. For large exchanges like Binance, deposits and withdrawals are typically free, while trades incur a fee of approximately 0.1% (with discounts for high-volume traders). Smaller exchanges may charge more, but discounts for large traders and market makers are common. Step 4 incurs FX conversion fees, often exceeding 1% for retail customers. Blockchain transfers also incur "gas" fees, which typically range from \$1–\$10 USD for ETH transactions (varies by network congestion and blockchain used). Since gas fees are fixed per transaction, transferring larger amounts is more cost-efficient.

Delays. Arbitrage transactions incur price risk, due to the time required for fund movements. For a regular customer, the entire process of moving funds across crypto exchanges may take days. Bank transfers, such as ACH transfers in the U.S., can take 1–3 business days. Blockchain transactions are much faster—ETH transfers typically settle within a minute. However, some exchanges impose a withdrawal hold on recently purchased crypto as a security measure. Selling ETH for KRW, withdrawing KRW, and converting it to USD may involve bank transfer delays similar to step 1.

Frequent arbitrageurs can mitigate price risk by simultaneously selling ETH on the Korean exchange at the time of purchase in the U.S. This effectively locks in profits before transfers settle. However, if short-selling is unavailable, this requires maintaining ETH and fiat balances on both exchanges.

Restrictions. In general, there are various restrictions and difficulties imposed by regulators on crypto firms attempting to use fiat infrastructure. Crypto firms face difficulty securing banking relationships in many jurisdictions.¹⁵ A report by the Alternative Investment Management Association states that 75% of surveyed crypto hedge funds had issues accessing or growing banking services for their funds, and an article by Bits about Money surveys various difficulties that crypto-related firms face accessing banking services in the US.

Inter-exchange arbitrage in particular faces additional restrictions. Regulators also impose KYC/AML requirements on entities interacting with crypto exchanges. For example, in the U.S., crypto exchanges are classified as money services businesses under the Bank Secrecy Act (1970) and must enforce strict KYC/AML policies.¹⁶ Capital controls also restrict arbitrage in some jurisdictions. For example, strict capital controls in South Korea limit the ability to withdraw KRW profits, making cross-border arbitrage difficult.¹⁷ Similarly, in India, crypto traders have received Foreign Exchange Management Act (FEMA) notices for offshore transfers, complicating arbitrage.¹⁸

¹⁵For example, Axios reports on "Operation Choke Point" in the U.S., which has made it harder for crypto firms to access banking services. Similarly, Forbes highlights the challenges blockchain companies face in securing banking relationships in the UK.

¹⁶See, for example, the 2019 joint statement by the CFTC, FinCEN, and the SEC.

¹⁷See Reuters.

 $^{^{18}\}mathrm{See}$ MNP Law.

B Analytical Solutions for Example Networks

B.1 Two-Node Network

Consider an economy with two exchanges, indexed by 1 and 2, and a single arbitrageur. From Definition (1), the equilibrium equations are:

$$p_{1} - \psi = -\gamma (x_{1} - z_{12})$$

$$p_{2} - \psi = -\gamma (x_{2} + z_{12})$$

$$\zeta z_{12} = p_{2} - p_{1}$$

The unique solution is:

$$p_{1} - \psi = -\gamma \left[\frac{(\zeta + \gamma) x_{1} + \gamma x_{2}}{2\gamma + \zeta} \right]$$
$$p_{2} - \psi = -\gamma \left[\frac{\gamma x_{1} + (\zeta + \gamma) x_{2}}{2\gamma + \zeta} \right]$$
$$z_{12} = \frac{\gamma}{2\gamma + \zeta} (x_{1} - x_{2})$$

B.2 Star Network

Consider a setting with N + 1 exchanges arranged in a star configuration. Let C denote the "central" exchange, and let $i = 1 \dots N$ index the remaining exchanges. From Definition (1), the equilibrium equations are:

$$p_C - \psi = -\gamma \left(x_C - \sum_{i=1}^N z_{C \to i} \right)$$
(28)

$$p_i - \psi = -\gamma \left(x_i + z_{C \to i} \right) \tag{29}$$

$$\zeta z_{iC} = p_C - p_i \tag{30}$$

The unique equilibrium equations are:

$$p_C - \psi = -\gamma \left[\frac{(\gamma + \zeta) x_C + \gamma \sum_{i=1}^N x_i}{\gamma (N+1) + \zeta} \right]$$
(31)

$$p_{i} - \psi = -\gamma \left[\frac{(\gamma^{2} + (1+N)\gamma\zeta + \zeta^{2})x_{i} + (\gamma^{2} + \gamma\zeta)x_{C} + \sum_{j \neq i}\gamma^{2}x_{j}}{(\gamma + \zeta)(\gamma(N+1) + \zeta)} \right]$$
(32)

$$z_{iC} = \frac{(N\gamma^2 + \gamma\zeta) x_i - (\gamma^2 + \gamma\zeta) x_C - \sum_{j \neq i} \gamma^2 x_j}{(\gamma + \zeta) (\gamma (N+1) + \zeta)}$$
(33)

There are many ways to arrive at (31) to (33); we briefly illustrate one method. First, consider a shock to the central node, so $x_C \neq 0$ and all $x_i = 0$. We can then easily solve (28) to (30), finding:

$$p_C - \psi = -\gamma \frac{\gamma + \zeta}{(\zeta + \gamma (N+1))} x_C$$
$$p_i - \psi = -\gamma \frac{\gamma}{(\zeta + \gamma (N+1))} x_C$$
$$z_{Ci} = \frac{\gamma}{\zeta + \gamma (N+1)} x_C$$

Now, suppose $x_1 \neq 0$, and suppose $x_C = 0$ and $x_j = 0$ for all $j \neq 1$. While slightly more involved, we can solve (28) to (30), finding:

$$p_{1} - \psi = -\gamma \left[\frac{(\gamma^{2} + (1+N)\gamma\zeta + \zeta^{2})x_{1}}{\gamma^{2} + N\gamma^{2} + 2\gamma\zeta + N\gamma\zeta + \zeta^{2}} \right]$$
$$p_{C} - \psi = -\gamma \left[\frac{(\gamma^{2} + \gamma\zeta)x_{1}}{\gamma^{2} + N\gamma^{2} + 2\gamma\zeta + N\gamma\zeta + \zeta^{2}} \right]$$
$$p_{i} - \psi = -\gamma \left[\frac{\gamma^{2}x_{1}}{\gamma^{2} + N\gamma^{2} + 2\gamma\zeta + N\gamma\zeta + \zeta^{2}} \right] \quad \forall j \neq i$$
$$z_{1C} = \frac{(N\gamma^{2} + \gamma\zeta)x_{1}}{\gamma^{2} + N\gamma^{2} + 2\gamma\zeta + N\gamma\zeta + \zeta^{2}}$$
$$z_{Cj} = \frac{\gamma^{2}x_{1}}{\gamma^{2} + N\gamma^{2} + 2\gamma\zeta + N\gamma\zeta + \zeta^{2}} \quad \forall j \neq i$$

Now, by symmetry, the effect of x_1 is identical to the effects of $x_2 \dots x_N$. Moreover, the structure of the equilibrium equations, (12), implies that solutions are linear in $x_C, x_1 \dots x_N$; thus, the effect of a combination of $x_C, x_1 \dots x_N$ is simply the sum of the effects of each individual shock. We can thus use the coefficients from the solutions to $x_C \neq 0$ and $x_1 \neq 0$ to fill in all coefficients in the equations (31) to (33).

C Proof of Proposition 1

Note that the variable ψ in Definition 1 simply shifts all equilibrium prices upwards and downwards; thus, for notational convenience, we will set $\psi = 0$ throughout this appendix; the expressions for $\psi \neq 0$ can be derived simply by replacing all price expressions with $p_i - \psi$.

C.1 Equilibrium Equations

ζ

The four strategies in Figure 2 represent a "cluster" of either one or two nodes, connected to the center of a star network, where there are many peripheral nodes not involved in the four colored strategies depicted in the bottom left of the figure. With slight abuse of notation, let N represent the number of peripheral exchanges in the star network: note that this N differs from the number of peripheral exchanges in the main text, since we have not counted the center node and the connected cluster of 1 or 2 nodes. As Figure 2 shows, we use C to refer to the network center; B is a node connected to C; and A is a node connected either to B in the green strategy, or to B and C in the orange strategy.

From Definition 1, the equilibrium equations for this system are first the equilibrium equations for the star network:

$$p_C = -\gamma \left(x_C + z_{BC} - \sum_{i=1}^N z_{Ci} \right) \tag{34}$$

$$p_i = -\gamma \left(x_i + z_{Ci} \right) \tag{35}$$

$$\zeta z_{Ci} = \qquad \qquad p_i - p_C \tag{36}$$

together with the equilibrium equations for the connected cluster. In the "blue" case of a single connected node B, we have:

$$p_B = -\gamma \left(x_B - z_{BC} \right) \tag{37}$$

$$z_{BC} = p_C - p_B \tag{38}$$

In the "red" case of two arbitrageurs $a \in \{1, 2\}$ and a single peripheral exchange B, we have:

$$p_B = -\gamma \left(x_B - z_{BC}^1 - z_{BC}^2 \right) \tag{39}$$

$$\zeta z_{BC}^1 = \qquad \qquad p_C - p_B \tag{40}$$

$$\zeta z_{BC}^2 = \qquad \qquad p_C - p_B \tag{41}$$

In the "green" case of two peripheral nodes, where A is connected to B and B to the center C, these are:

$$p_B = -\gamma \left(x_B + z_{AB} - z_{BC} \right) \tag{42}$$

$$p_A = -\gamma \left(x_A - z_{AB} \right) \tag{43}$$

$$\zeta z_{BC} = p_C - p_B \tag{44}$$

$$\zeta z_{AB} = \qquad \qquad p_B - p_A \tag{45}$$

In the "orange" case of two peripheral nodes A, B both connected to each other and the center C, we have:

$$p_B = -\gamma \left(x_B + z_{AB} - z_{BC} \right) \tag{46}$$

$$p_A = -\gamma \left(x_A - z_{AB} - z_{AC} \right) \tag{47}$$

$$\zeta z_{BC} = \qquad \qquad p_C - p_B \tag{48}$$

$$\zeta z_{AB} = \qquad \qquad p_B - p_A \tag{49}$$

$$\zeta z_{AC} = \qquad \qquad p_C - p_A \tag{50}$$

C.2 Convergence to the $N = \infty$ Limit

We will now characterize the limit as the number of peripheral exchanges $N \to \infty$ for the "green" strategy; arguments for the other strategies are analogous. First, note that the equations (34) to (36) are exactly the equilibrium conditions for the star network in Appendix B.2, except replacing x_C with $x_C + z_{BC}$. From Appendix B.2, the unique solution for p_C is thus:

$$p_C = -\gamma \left[\frac{(\gamma + \zeta) \left(x_C + z_{BC} \right) + \gamma \sum_{i=1}^N x_i}{\gamma \left(N + 1 \right) + \zeta} \right]$$
(51)

We can thus write (51) as:

$$p_C = y_C^N + \xi z_{BC}^N \tag{52}$$

where,

$$\xi^N \equiv -\gamma \frac{\gamma + \zeta}{\gamma \left(N+1\right) + \zeta} \tag{53}$$

$$y_C^N \equiv -\gamma \frac{(\gamma + \zeta) x_C + \gamma \sum_{i=1}^N x_i}{\gamma (N+1) + \zeta}$$
(54)

Intuitively, y_C^N is a random variable, which is a function of $x_1 \dots x_N$ and x_C ; it behaves like an "intercept" or "constant" term for p_C , in the sense that we have $p_C = y_C$ when $z_{BC} = 0$. The

constant ξ^N is a slope term, capturing the effect of z_{BC} on p_C . A set of prices p_A, p_B, p_C and transfers z_{AB}, z_{BC} thus forms an equilibrium under shocks $x_1 \dots x_N, x_C, x_B, x_A$ if and only if the prices and transfers satisfy (52), in addition to the A, B cluster equilibrium equations (42) to (45); for such a system, peripheral prices p_i and flows z_{iC} within the star network can be solved for using expressions (32) and (33).

An intuition behind (52) is that, when the A, B cluster is connected to the star network via node C, node C in fact behaves like a single node, with a different "inventory shock" y_C^N , which is an aggregate of all the star network inventory shocks, and a different inventory responsiveness ξ , reflecting the fact that inventory sold in node C will be redistributed via arbitrage to all peripheral nodes.

We can thus write the solution of the A, B cluster system, for any N, as the 5-equation system:

$$p_B + \gamma z_{AB} - \gamma z_{BC} = -\gamma x_B \tag{55}$$

$$p_A - \gamma z_{AB} = -\gamma x_A \tag{56}$$

$$\zeta z_{AB} - p_B + p_A = 0 \tag{58}$$

$$p_C - \xi^N z_{BC} = \qquad \qquad y_C^N \tag{59}$$

Letting \boldsymbol{M}^N denote the matrix representation of the system on the left, we have:

$$\begin{pmatrix} p_B \\ p_A \\ \zeta z_{BC} \\ \zeta z_{AB} \\ p_C \end{pmatrix} = \left(\boldsymbol{M}^N \right)^{-1} \begin{pmatrix} -\gamma x_B \\ -\gamma x_A \\ 0 \\ 0 \\ y_C^N \end{pmatrix}$$
(60)

where \boldsymbol{M}^{N} is a series of constant matrices – with dependence on N because ξ^{N} is an element of \boldsymbol{M}^{N} – and y_{C}^{N} is a series of random variables, both indexed by the number of peripheral exchanges N.

Now, we take the limit as $N \to \infty$. Notice that we can write (54) as:

$$y_C^N \equiv -\gamma \left(\frac{(\gamma + \zeta) x_C}{\gamma (N+1) + \zeta} + \frac{\gamma \sum_{i=1}^N x_i}{\gamma (N+1) + \zeta} \right)$$
(61)

We assumed that $x_1 \ldots x_N, x_C$ are i.i.d. and bounded; hence, the weak law of large numbers

(WLLN) implies that the sequence y_C^N converges to the constant 0 in probability as $N \to \infty$, and moreover y_C^N is uniformly integrable. Similarly, (53) shows that the constant ξ^N converges to 0 as $N \to \infty$.

We then apply the continuous mapping theorem to (60). We will show in Appendix C.3 below that (60) is solvable at $y_C^N = 0, \xi^N = 0$, by explicitly solving the system. This implies that the solution of (60) is a continuous and differentiable function jointly of y_C^N , which is an element of the vector on the RHS, and ξ^N , which is an element of the \mathbf{M}^N matrix; formally, this is because the function $\mathbf{M}^N \to (\mathbf{M}^N)^{-1}$ is continuously differentiable wherever \mathbf{M}^N is invertible, and the matrix-vector multiplication on the RHS of (60) is also a smooth operation.

The continuous mapping theorem immediately implies that, as $N \to \infty$, all outcomes $p_B, p_A, z_{BC}, z_{AB}, p_C$ converge in probability to their values in the limiting case, which we can write as:

$$y_C^\infty = 0, \xi^\infty = 0$$

We additionally require convergence of second moments of outcome variables, in order to calculate arbitrageur surplus. To show this, first note that $(y_C^N)^2$ is uniformly integrable (U.I.) by the WLLN for bounded random variables x_i , and $(\xi^N)^2$ is trivially U.I. since it is a constant. Second, the mapping in (60) is differentiable at the limit point. Thus, since the LHS of (60) is an asymptotically linear function of the U.I. sequences of random variables $(y_C^N)^2$ and $(\xi^N)^2$, the second moments of all outcomes on the LHS of (60) also converge to their limit values as $N \to \infty$.

The limit of the sequence of outcomes has a particularly simple representation because (59) simply reduces to the condition that $p_C = 0$. Intuitively, this limit case describes equilibrium in a scenario where the A, B cluster is connected to an infinitely deep market, where an arbitrary amount can be bought and sold at price 0. We next characterize equilibrium outcomes in this limit case.

C.3 Characterizing Equilibria in the $N \to \infty$ Limit

Now, we characterize outcomes for the four colored strategies in the limit case of $p_C = 0$ (recall that we normalized $\psi = 0$, so in general in the limit $p_C = \psi$).

C.3.1 Blue: One Node

First, we consider the blue strategy in Figure 2: consider a single peripheral exchange B connected to the central exchange C by a single arbitrageur. The equilibrium equations are (37) and (38), setting $p_C = 0$. The solutions are:

$$p_B = -\gamma \left(\frac{\zeta}{\gamma + \zeta}\right) x_B \tag{62}$$

$$z_{BC} = \frac{\gamma}{\zeta + \gamma} x_B \tag{63}$$

From (6), expected arbitrageur profits are proportional to the expectation of $(p_B - \psi)^2$; from (62), this is just

$$\gamma^2 \left(\frac{\zeta}{\gamma+\zeta}\right)^2 \sigma^2 \tag{64}$$

where σ^2 is the variance of x_B . Multiplying by $\frac{1}{2\zeta}$, we get the expected profit of the blue strategy.

C.3.2 Red: One Node, Two Arbitrageurs

Now we consider the red strategy, characterized by equations (39) to (41): suppose there are two arbitrageurs, $a \in \{1, 2\}$, connecting a single peripheral exchange B to the central exchange C. Solving the system, equilibrium prices are:

$$p_B = -\gamma \left(\frac{\zeta}{2\gamma + \zeta}\right) x_B \tag{65}$$

Arbitrageurs' profits are again proportional to the expectation of $(p_B - \psi)^2$, which is:

$$\gamma^2 \left(\frac{\zeta}{2\gamma + \zeta}\right)^2 \sigma^2 \tag{66}$$

Multiplying by $\frac{1}{2\zeta}$, we get the expected profit of the red strategy. Now, (66) is strictly lower than (64); thus, the red strategy in Figure 2 is always dominated by the blue strategy in the limit as $N \to \infty$, and we need not consider equilibria in which the red strategy is played.

C.3.3 Green: Two Nodes, One Central Link

Next, we consider the green strategy, characterized by (42) to (45), setting $p_C = 0$: suppose there are two peripheral exchanges A, B, such that A is connected to B, and B is connected to the central exchange C. The solutions are:

$$p_{A} = -\gamma \left[\frac{(2\gamma\zeta + \zeta^{2}) x_{A} + \gamma\zeta x_{B}}{\gamma^{2} + 3\gamma\zeta + \zeta^{2}} \right]$$

$$p_{B} = -\gamma \left[\frac{\gamma\zeta x_{A} + (\gamma\zeta + \zeta^{2}) x_{B}}{\gamma^{2} + 3\gamma\zeta + \zeta^{2}} \right]$$

$$z_{BC} = \frac{\gamma^{2} x_{A} + (\gamma^{2} + \gamma\zeta) x_{B}}{\gamma^{2} + 3\gamma\zeta + \zeta^{2}}$$

$$z_{AB} = \frac{(\gamma^{2} + \gamma\zeta) x_{A} - \gamma\zeta x_{B}}{\gamma^{2} + 3\gamma\zeta + \zeta^{2}}$$

To characterize the green strategy's profits, we are interested in the profit of the A, B arbitrageur. This is:

$$p_A - p_B = \frac{\gamma \zeta \left(\zeta x_B - \left(\zeta + \gamma\right) x_A\right)}{\gamma^2 + 3\gamma \zeta + \zeta^2}$$

In expectation, with x_B, x_A independent with variance σ^2 , this is:

$$E\left[\left(p_A - p_B\right)^2\right] = \frac{\gamma^2 \zeta^2 \left(\gamma^2 + 2\gamma\zeta + 2\zeta^2\right)}{\left(\gamma^2 + 3\gamma\zeta + \zeta^2\right)^2} \sigma^2 \tag{67}$$

Multiplying by $\frac{1}{2\zeta}$, we get the expected profit of the green strategy in Figure 2.

C.3.4 Orange: Two Nodes, Two Central Links

Finally, we consider the orange strategy, characterized by equations (46) to (50), setting $p_C = 0$. The solutions are:

$$p_{A} = -\gamma \left[\frac{(2\gamma\zeta + \zeta^{2}) x_{A} + \gamma\zeta x_{B}}{3\gamma^{2} + 4\gamma\zeta + \zeta^{2}} \right]$$

$$p_{B} = -\gamma \left[\frac{\gamma\zeta x_{A} + (2\gamma\zeta + \zeta^{2}) x_{B}}{3\gamma^{2} + 4\gamma\zeta + \zeta^{2}} \right]$$

$$z_{BC} = \frac{\gamma^{2} x_{A} + (2\gamma^{2} + \gamma\zeta) x_{B}}{3\gamma^{2} + 4\gamma\zeta + \zeta^{2}}$$

$$z_{AB} = \frac{\gamma (x_{A} - x_{B})}{3\gamma + \zeta}$$

$$z_{AC} = \frac{(2\gamma^{2} + \gamma\zeta) x_{A} + \gamma^{2} x_{B}}{3\gamma^{2} + 4\gamma\zeta + \zeta^{2}}$$

The A, B price gap is then:

$$p_A - p_B = \frac{\left(x_B - x_A\right)\gamma\zeta}{3\gamma + \zeta}$$

Arbitrageurs' payoffs are $\frac{1}{2\zeta}$ times the variance of this price gap, which is:

$$E\left[\left(p_A - p_B\right)^2\right] = \frac{2\gamma^2 \zeta^2}{\left(3\gamma + \zeta\right)^2} \sigma^2 \tag{68}$$

The difference between the price gap under the green strategy, (67), and the price gap under the orange strategy, (68), rearranges to:

$$\frac{\gamma^{3}\zeta^{2}\left(7\gamma^{3}+12\gamma^{2}\zeta+9\gamma\zeta^{2}+2\zeta^{3}\right)}{\left(3\gamma+\zeta\right)^{2}\left(\gamma^{2}+3\gamma\zeta+\zeta^{2}\right)^{2}}\sigma^{2}$$

Since $\gamma > 0$ and $\zeta > 0$, this difference is always positive. Thus, regardless of ζ and γ , the green strategy always induces a larger variance in the p_A, p_B price gap, and thus is more profitable for the A, B arbitrageur in expectation, than the orange strategy. This demonstrates the economically intuitive fact that the orange strategy of linking two connected peripheral nodes is always dominated by the green strategy of linking an isolated node to a peripheral node.

C.3.5 Comparing Blue and Green Strategies

Since we have shown that the red strategy is dominated by the blue strategy, and the orange strategy is dominated by the green strategy, a star-network equilibrium exists for any parameters such that the blue strategy pays higher than the green strategy. Taking the ratio of (67) and (64), we require:

$$\frac{\left(\gamma+\zeta\right)^2\left(\gamma^2+2\gamma\zeta+2\zeta^2\right)}{\left(\gamma^2+3\gamma\zeta+\zeta^2\right)^2} < 1$$

Rearranging, this is:

$$(\gamma+\zeta)^2\left(\gamma^2+2\gamma\zeta+2\zeta^2\right)-\left(\gamma^2+3\gamma\zeta+\zeta^2\right)^2<0$$

Expanding and simplifying, we have:

$$\zeta^3 - 2\gamma^3 - 4\gamma^2\zeta < 0$$

Letting $\theta \equiv \frac{\zeta}{\gamma}$, this is:

$$\theta^3 - 4\theta - 2 < 0$$

Since $\zeta > 0, \gamma > 0$, we have $\theta > 0$. It can be verified that θ is negative on the interval [0, 2.21432...], and positive otherwise. Thus, the blue strategy delivers higher expected

utility than the green strategy whenever $\frac{\zeta}{\gamma} < 2.21432$.

C.4 Completing the Proof

Together, Appendix C.2 showed that, up to at least second moments, all equilibrium quantities in *B* or *A*, *B* modules converge to their values derived in Appendix C.3 as $N \to \infty$.¹⁹ Hence, it is possible to find *N* sufficiently large that expected arbitrageur profits – which depend only on second moments of equilibrium prices – in the red, green, orange, and blue strategies are arbitrarily close to these limit values. Appendix C.3 then showed that, in the limit, the red strategy is dominated by the blue strategy, the orange strategy is dominated by the green strategy, and the blue strategy has higher payoffs than the green strategy if and only if $\frac{\zeta}{\gamma} < 2.21432$. We have thus proved Proposition 1: whenever $\frac{\zeta}{\gamma} < 2.21432$, for sufficiently large *N*, the blue strategy's payoffs will be strictly higher than all other strategies' payoffs, so a star-shaped equilibrium exists.

 $^{^{19}}$ As a minor aside, note that with slight abuse of notation we used N in this appendix to represent the number of peripheral exchanges, excluding the central exchanges and the A, B cluster. This is either two or three exchanges less than the total number of exchanges, which we refer to as N in the main text. This does not affect our limit arguments.

D Price Impact

In the baseline model, we assumed for simplicity that arbitrageurs bid competitively, ignoring the impact of their behavior on market prices. A more natural assumption is that arbitrageurs bid strategically, optimally accounting for their own impact on prices. This appendix solves the model assuming strategic bidding. We prove the following proposition:

Proposition 2. In the model with strategic bidding, a star network equilibrium exists whenever:

$$\frac{\zeta}{\gamma} < 1.07395$$

Proposition 2 is qualitatively similar to Proposition 1 in the main text: star-network equilibria exist when the ratio $\frac{\zeta}{\gamma}$ is sufficiently low. However, the cutoff on $\frac{\zeta}{\gamma}$ is slightly lower under strategic bidding than the cutoff in (13) in the main text. While the result is qualitatively similar, the derivations are substantially more complex; we thus focus on the competitive case in the main text for simplicity.

As in Appendix C, for notational simplicity we assume $\psi = 0$ throughout.

Suppose an arbitrageur a links nodes i, j within some network. The arbitrageur can be thought of facing a random *inverse residual supply* function, specifying how the price gap

$$\Delta p_{ji} = p_j - p_i$$

depends on her trade quantity z_{ij} . In general, this inverse residual supply function has the form:

$$\Delta p_{ji} = -Kz_{ij} + \eta_{ij}$$

where K > 0 is a slope parameter, and η_{ij} is a random intercept term. In our setting, since randomness results from zero-mean inventory shocks x_{ij} , η_{ij} will also have mean 0 in our setting.

Following the literature on demand-function submission games (Klemperer and Meyer, 1989; Rostek and Yoon, 2020), we require the arbitrageur to submit a *demand schedule*, specifying z_{ij} (Δp_{ji}), the amount she is willing to transport between *i* and *j* as a function of the price gap Δp_{ji} . We require the demand schedule to be an ex-post best response, implementing the arbitrageur's optimal point on her residual supply curve for every realization of η_{ij} . We prove the following proposition, characterizing the arbitrageur's best response, in Appendix D.6 below. **Proposition 3.** Suppose an arbitrageur faces an inverse residual supply schedule of the form:

$$\Delta p_{ji} = -K z_{ij} + \eta_{ij} \tag{69}$$

where η_{ij} has mean 0. The arbitrageur's best response is to submit the demand schedule:

$$z_{ij}\left(\Delta p_{ji}\right) = \frac{\Delta p_{ji}}{K + \zeta} \tag{70}$$

The arbitrageur's expected utility is:

$$\frac{Var\left[\eta_{ij}\right]}{4K+2\zeta}\tag{71}$$

Note that, if the arbitrageur bids competitively, setting price equal to marginal cost and ignoring her price impact, from (5), her demand schedule satisfies:

$$z_{ij}\left(\Delta p_{ji}\right) = \frac{\Delta p_{ji}}{\zeta} \tag{72}$$

Hence, as is known in the literature, strategic arbitrageurs shade bids: the slope of (70) with respect to Δp_{ji} is lower than the slope of (72). Arbitrageurs shade bids more the larger K, the slope of inverse residual supply, is; that is, arbitrageurs shade more when withholding supply has a greater impact on market prices. The arbitrageur's final expected utility, (71), is a function of the variance in the intercept parameter η_{ij}^2 , the slope of inverse residual supply K, and the arbitrageur's transportation cost parameter ζ .

We will demonstrate conditions under which star network equilibria exist under strategic bidding, in the limit as $N \to \infty$, analogously to the proof of Proposition 1. We will characterize the asymptotic expected payoffs of the red, green, orange, and blue strategies in Figure 2, in the limit as the central node becomes an infinitely elastic source at a fixed price $p_C = 0$. We will then find conditions on ζ, γ under which the blue strategy of completing the star pays higher than the red and green strategies.

D.1 Blue: One Node

As in appendix C.3.1, we consider the "blue" strategy, where an arbitrageur links an isolated node B to the center C. From (37), customers' demand is:

$$p_B = -\gamma \left(x_B - z_{BC} \right)$$

Now the price at the center is $p_C = 0$, so we immediately have:

$$\Delta p_{C,B} = -p_B = \gamma x_B - \gamma z_{BC} \tag{73}$$

We can then apply Proposition 3 with $K = \gamma$, to find that the arbitrageur's optimal bid is:

$$z_{BC} = \frac{\Delta p_{CB}}{\gamma + \zeta} \tag{74}$$

Now, in (73), the intercept term is $\eta_{BC} = \gamma x_B$, so:

$$Var\left[\eta_{BC}\right] = Var\left(\gamma x_B\right) = \gamma^2 \sigma^2$$

Hence, from (71), arbitrageur expected utility is:

$$\frac{\gamma^2}{4\gamma + 2\zeta}\sigma^2\tag{75}$$

D.2 Red: One Node, Two Arbitrageurs

Next, we consider the payoffs of the "red" strategy, where two arbitrageurs link a single exchange to the center. Analogous to (39), customers' demand is:

$$p_B = -\gamma \left(x_B - z_{BC}^1 - z_{BC}^2 \right) \tag{76}$$

The difficulty in directly applying Proposition 3 is that the residual supply curve one arbitrageur faces depends on the other arbitrageurs' bid slope. Thus, we need to find a symmetric equilibrium in bid slopes, in which each arbitrageur's chosen bid slope is a best response to the residual supply induced by the other arbitrageur's bid slope. Conjecture a symmetric equilibrium bid curve:

$$z_{BC} = z_{BC}^1 = z_{BC}^2 = b_{BC} \left(-p_B\right) \tag{77}$$

Substituting (77) into (76) for z_{BC}^1 , we have from arbitrageur 2's perspective:

$$p_B = -\gamma \left(x_B + b_{BC} p_B - z_{BC} \right)$$

Solving for p_B , we have:

$$p_B = \frac{-\gamma x_B}{1 + b_{BC}\gamma} + \frac{\gamma z_{BC}}{1 + b_{BC}\gamma}$$

Now since the price at the center is $p_C = 0$, we immediately have:

$$\Delta p_{C,B} = -p_B = \frac{\gamma x_B}{1 + b_{BC}\gamma} - \frac{\gamma z_{BC}}{1 + b_{BC}\gamma} \tag{78}$$

We can then apply (70) of Proposition 3, taking the z_{BC} coefficient in (78):

$$K_{BC} = \frac{\gamma}{1 + b_{BC}\gamma} \tag{79}$$

Plugging into (70), if one arbitrageur's bid slope is b_{BC} , the other arbitrageur's optimal bid curve is:

$$z_{BC} = \frac{\Delta p_{CB}}{\frac{\gamma}{1+b_{BC}\gamma} + \zeta} \tag{80}$$

Now, b_{BC} in (77) was defined as the coefficient on price in the submitted demand curve. Thus, in equilibrium we must have that (80) is consistent with (77):

$$b_{BC} = \frac{1}{\frac{\gamma}{1+b_{BC}\gamma} + \zeta} \tag{81}$$

Note that b_{BC} in (81) is always greater than $\frac{1}{\gamma+\zeta}$, the bid slope in the case with one arbitrageur and one node, in (74). This is because the arbitrageur faces an elastic demand curve from the other arbitrageur, thus increasing the elasticity of residual supply, causing her to also bid more steeply. (81) defines a quadratic equation, whose unique positive solution is:

$$b_{BC} = \frac{\sqrt{\zeta}\sqrt{4\gamma + \zeta} - \zeta}{2\gamma\zeta} \tag{82}$$

This is thus the unique equilibrium bid slope. To evaluate expected profits, note that from (78), we have:

$$\eta_{BC} = \frac{\gamma}{1 + b_{BC}\gamma} x_B$$

Hence,

$$Var\left[\eta_{BC}\right] = \left(\frac{\gamma}{1 + b_{BC}\gamma}\right)^2 \sigma^2 \tag{83}$$

We can then plug in b_{BC} from (82) into (83), and we can use the value of K_{BC} from (79), to get expected profits:

$$\frac{\gamma^2}{\left(\frac{\sqrt{\zeta}\sqrt{4\gamma+\zeta}-\zeta}{2\zeta}+1\right)^2 \left(\frac{4\gamma}{\frac{\sqrt{\zeta}\sqrt{4\gamma+\zeta}-\zeta}{2\zeta}+1}+2\zeta\right)}\sigma^2$$

Rearranging somewhat, we have:

$$\frac{\gamma^2}{\left(4\gamma\left(\frac{\sqrt{\zeta}\sqrt{4\gamma+\zeta}-\zeta}{2\zeta}+1\right)+2\zeta\left(\frac{\sqrt{\zeta}\sqrt{4\gamma+\zeta}-\zeta}{2\zeta}+1\right)^2\right)}\sigma^2\tag{84}$$

Now,

$$\frac{\sqrt{\zeta}\sqrt{4\gamma+\zeta}-\zeta}{2\zeta}+1\geq 1$$

which implies that the red strategy's expected profits, (84), are always less then or equal to the blue strategy's expected profits, (75). Thus, as in the baseline model, the red strategy of competing with an arbitrageur over a peripheral exchange remains dominated by the blue strategy of connecting to a new peripheral exchange, under strategic bidding.

D.3 Green: Two Nodes, One Central Link

Next, we consider the payoff to the green strategy, of linking one peripheral node to another peripheral node. Here, once again, the difficulty in applying Proposition 3 is that the residual supply curve facing one arbitrageur depends on the other arbitrageur's bid slope; thus, again, we need to find an equilibrium in bid slopes. In such an equilibrium, bids will take the form:

$$z_{BC} = b_{BC} \left(-p_B\right) \tag{85}$$

$$z_{AB} = b_{AB} \left(p_B - p_A \right) \tag{86}$$

for some coefficients b_{BC} , b_{AB} , which must be equal to the RHS of (70) in equilibrium. Note that these two equations resemble (44) and (45), but with different coefficients on prices. The customer demand functions remain unchanged from (42) and (43):

$$p_B = -\gamma \left(x_B + z_{AB} - z_{BC} \right) \tag{87}$$

$$p_A = -\gamma \left(x_A - z_{AB} \right) \tag{88}$$

We now consider the best-responses of the B, C and A, B arbitrageurs respectively.

D.3.1 B, C Arbitrageur

The B, C arbitrageur faces a residual supply curve which results from the combined behavior of customers at both links and the A, B arbitrageur. To solve for this, take equations (86),

(88), (87) and note that we can solve for p_B , eliminating p_A and z_{AB} . This gives:

$$p_B = \frac{1}{2b_{AB}\gamma + 1} \left(-b_{AB}\gamma^2 x_A - \left(\gamma + b_{AB}\gamma^2\right) x_B + \left(b_{AB}\gamma^2 + \gamma\right) z_{BC} \right)$$

Thus, using that $p_C = 0$, we have:

$$\Delta p_{CB} = p_C - p_B = -p_B = -\frac{1}{2b_{AB}\gamma + 1} \left(-b_{AB}\gamma^2 x_A - \left(\gamma + b_{AB}\gamma^2\right) x_B + \left(b_{AB}\gamma^2 + \gamma\right) z_{BC} \right)$$
(89)

We can then apply (70) of Proposition 3, taking the z_{BC} coefficient in (89):

$$K_{BC} = \frac{b_{AB}\gamma^2 + \gamma}{2b_{AB}\gamma + 1}$$

to calculate the B, C arbitrageur's optimal bid curve:

$$z_{BC}\left(\Delta p_{CB}\right) = \frac{\Delta p_{CB}}{K+\zeta} = \frac{\Delta p_{CB}}{\frac{b_{AB}\gamma^2 + \gamma}{2b_{AB}\gamma + 1} + \zeta}$$

Since we defined b_{BC} as the price coefficient in B, C's optimal demand schedule, we can then write B, C's best response bid slope, as a function of b_{AB} , as simply:

$$b_{BC} = \frac{1}{\frac{b_{AB}\gamma^2 + \gamma}{2b_{AB}\gamma + 1} + \zeta} \tag{90}$$

Note that we can write:

$$b_{BC} = \frac{1}{\frac{b_{AB}\gamma + 1}{2b_{AB}\gamma + 1}\gamma + \zeta}$$

Thus, we will always have:

$$b_{BC} > \frac{1}{\gamma + \zeta}$$

which is the bid slope of the one-node arbitrageur, from (74) of Subsection D.1. Intuitively, the B, C arbitrageur in this setting faces slightly more elastic demand compared to Subsection D.1, due to the fact that node B is connected to node A; she thus bids slightly more aggressively.

D.3.2 A, B Arbitrageur

Similarly, the A, B arbitrageur faces a residual supply curve which results from the behavior of customers and the B, C arbitrageur. To solve for this, we can first solve (85) and (87) to find:

$$p_B = \frac{-\gamma x_B - \gamma z_{AB}}{1 + b_{BC}\gamma}$$

Thus, using (88),

$$\Delta p_{BA} = p_B - p_A = \frac{-\gamma x_B - \gamma z_{AB}}{1 + b_{BC} \gamma} - \left[-\gamma \left(x_A - z_{AB}\right)\right]$$
$$\Delta p_{BA} = \gamma x_A - \frac{\gamma}{1 + b_{BC} \gamma} x_B - z_{AB} \left(\gamma + \frac{\gamma}{1 + b_{BC} \gamma}\right) \tag{91}$$

We can thus apply (70) of Proposition 3, taking the z_{AB} coefficient in (91):

$$K_{AB} = \gamma + \frac{\gamma}{1 + b_{BC}\gamma} \tag{92}$$

Hence, the A, B arbitrageur's best response condition is:

$$b_{AB} = \frac{1}{\left(\gamma + \frac{\gamma}{1 + b_{BC}\gamma}\right) + \zeta} \tag{93}$$

Notice that (93) will always be slightly lower than $\frac{1}{\gamma+\zeta}$, the bid slope of the one-node arbitrageur in Subsection D.1. Intuitively, the A, B arbitrageur connects an isolated node to an imperfectly elastic peripheral node; she always faces slightly more price impact than an arbitrageur connecting to a perfectly elastic center, and thus shades bids more. As $b_{BC} \to \infty$, so the peripheral node's demand slope becomes perfectly elastic, the AB arbitrageur's best-response bid slope converges to $\frac{1}{\gamma+\zeta}$.

D.3.3 Solving for Equilibrium

An equilibrium in the two-node setting requires expressions (90) and (93) to be satisfied simultaneously. These two equations are quadratic, and have two pairs of solutions, in which only one pair is positive; the positive solutions are:

$$b_{AB} = \frac{\sqrt{32\gamma^4 + 72\gamma^3\zeta + 48\gamma^2\zeta^2 + 12\gamma\zeta^3 + \zeta^4 - 2\gamma\zeta - \zeta^2}}{2(4\gamma^3 + 7\gamma^2\zeta + 2\gamma\zeta^2)}$$
(94)
$$b_{BC} = \frac{1}{2\gamma^{2} + 4\gamma\zeta + \zeta^{2}} \left[-\frac{9\gamma^{2}\zeta^{2}}{4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2}} - \frac{4\gamma^{3}\zeta}{4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2}} - \frac{\zeta^{4}}{4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2}} - \frac{\zeta^{4}}{4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2}} - \frac{11\gamma\zeta^{3}}{2(4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2})} + \frac{\zeta^{2}\sqrt{32\gamma^{4} + 72\gamma^{3}\zeta + 48\gamma^{2}\zeta^{2} + 12\gamma\zeta^{3} + \zeta^{4}}}{4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2}} + \frac{7\gamma\zeta\sqrt{32\gamma^{4} + 72\gamma^{3}\zeta + 48\gamma^{2}\zeta^{2} + 12\gamma\zeta^{3} + \zeta^{4}}}{2(4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2})} + \frac{2\gamma^{2}\sqrt{32\gamma^{4} + 72\gamma^{3}\zeta + 48\gamma^{2}\zeta^{2} + 12\gamma\zeta^{3} + \zeta^{4}}}{4\gamma^{3} + 7\gamma^{2}\zeta + 2\gamma\zeta^{2}} - \zeta \right]$$
(95)

To evaluate the payoff to the green strategy, we wish to to evaluate the expected profit of the A, B arbitrageur in equilibrium. We can calculate the equilibrium value of K_{AB} by plugging the expression for b_{BC} from (95) into (92). Moreover, from (91), we have:

$$\eta_{AB} = \gamma x_A - \frac{\gamma}{1 + b_{BC} \gamma} x_B$$

Hence, using that we assumed inventory shocks are independent,

$$Var\left[\eta_{AB}\right] = \left(\gamma^2 + \left(\frac{\gamma}{1 + b_{BC}\gamma}\right)^2\right)\sigma^2$$

where we can once again plug in the equilibrium b_{BC} value from (95). We can then use expression (71) of Proposition 3 to calculate the A, B arbitrageur's expected profits. These expressions are too convoluted to usefully display; we include them in a companion Mathematica notebook to the paper.

D.4 Orange: Two Nodes, Two Central Links

We now consider the orange strategy. Equilibrium bids are:

$$z_{BC} = b_C \left(-p_B\right) \tag{96}$$

$$z_{AC} = b_C \left(-p_A\right) \tag{97}$$

$$z_{AB} = b_{AB} \left(p_B - p_A \right) \tag{98}$$

where we used the symmetry-imposed restriction that b_C is common to the B, C and A, C arbitrageurs. The customer demand functions are:

$$p_B = -\gamma \left(x_B + z_{AB} - z_{BC} \right) \tag{99}$$

$$p_A = -\gamma \left(x_A - z_{AB} - z_{AC} \right) \tag{100}$$

We now consider the best responses of the B, C arbitrageur (thus also the A, C arbitrageur by symmetry) and the A, B arbitrageur.

D.4.1 B, C arbitrageur

The B, C arbitrageur faces a residual supply curve which results from the combined behavior of customers at both links and the A, B arbitrageur. To solve for this, take equations (97), (98), (99), (100) and note that we can solve for p_B , eliminating p_A, z_{AB} , and z_{AC} . This gives:

$$p_B = \frac{-b_{AB}\gamma^2 x_A - (\gamma + b_{AB}\gamma^2 + b_C\gamma^2) x_B + (\gamma + b_{AB}\gamma^2 + b_C\gamma^2) z_{BC}}{1 + 2b_{AB}\gamma + b_C\gamma + b_{AB}b_C\gamma^2}$$

Thus, using that $p_C = 0$, we have:

$$\Delta p_{CB} = p_C - p_B = -p_B = -\frac{-b_{AB}\gamma^2 x_A - (\gamma + b_{AB}\gamma^2 + b_C\gamma^2 x_B) + (\gamma + b_{AB}\gamma^2 + b_C\gamma^2) z_{BC}}{1 + 2b_{AB}\gamma + b_C\gamma + b_{AB}b_C\gamma^2}$$
(101)

We can then apply (70) of Proposition 3, taking the z_{BC} coefficient in (101):

$$K_{BC} = \frac{\gamma + b_{AB}\gamma^2 + b_C\gamma^2}{1 + 2b_{AB}\gamma + b_C\gamma + b_{AB}b_C\gamma^2}$$

Hence, the B, C arbitrageur's best response condition is:

$$b_C = \frac{1}{\left(\frac{\gamma + b_{AB}\gamma^2 + b_C\gamma^2}{1 + 2b_{AB}\gamma + b_C\gamma + b_{AB}b_C\gamma^2}\right) + \zeta}$$
(102)

Note that we can write:

$$b_{C} = \frac{1}{\left(\frac{1+b_{AB}\gamma+b_{C}\gamma}{1+2b_{AB}\gamma+b_{C}\gamma+b_{AB}b_{C}\gamma^{2}}\right)\gamma+\zeta}$$

so once again, we always have:

$$b_C > \frac{1}{\gamma + \zeta}$$

by a similar argument to Appendix D.3.1.

D.4.2 A, B arbitrageur

We solve the system defined by (96), (97), (99), and (100), finding that:

$$\Delta p_{BA} = p_B - p_A = \frac{1}{1 + b_C \gamma} \left(\gamma x_A - \gamma x_B - 2\gamma z_{AB} \right) \tag{103}$$

We can thus apply (70) of Proposition 3, taking the z_{AB} coefficient in (103):

$$K_{AB} = \frac{2\gamma}{1 + b_C \gamma} \tag{104}$$

Hence, the A, B arbitrageur's best response condition is:

$$b_{AB} = \frac{1}{\left(\frac{2\gamma}{1+b_C\gamma}\right) + \zeta} \tag{105}$$

Unlike Appendix D.3.2, (105) is not guaranteed to be lower than $\frac{1}{\gamma+\zeta}$.

D.4.3 Solving for Equilibrium

An equilibrium in the two-node setting requires expressions (102) and (105) to be satisfied simultaneously. These two equations are quadratic, and have two pairs of solutions, in which only one pair is both positive; the positive solutions are:

$$b_C = \frac{-2\gamma^2 + \sqrt{4\gamma^4 + 48\gamma^3\zeta + 44\gamma^2\zeta^2 + 12\gamma\zeta^3 + \zeta^4} - 4\gamma\zeta - \zeta^2}{2(2\gamma^2\zeta + \gamma\zeta^2)}$$
(106)

$$b_{AB} = \frac{1}{-4\gamma^2 + 2\gamma\zeta + \zeta^2} \left[\frac{6\gamma^2\zeta^2}{2\gamma^2\zeta + \gamma\zeta^2} + \frac{\gamma\zeta^3}{2\gamma^2\zeta + \gamma\zeta^2} + \frac{4\gamma^4}{2\gamma^2\zeta + \gamma\zeta^2} + \frac{4\gamma^4}{2\gamma^2\zeta + \gamma\zeta^2} + \frac{10\gamma^3\zeta}{2\gamma^2\zeta + \gamma\zeta^2} - \frac{2\gamma^2\sqrt{4\gamma^4 + 48\gamma^3\zeta + 44\gamma^2\zeta^2 + 12\gamma\zeta^3 + \zeta^4}}{2\gamma^2\zeta + \gamma\zeta^2} - \frac{\gamma\zeta\sqrt{4\gamma^4 + 48\gamma^3\zeta + 44\gamma^2\zeta^2 + 12\gamma\zeta^3 + \zeta^4}}{2\gamma^2\zeta + \gamma\zeta^2} - \frac{\gamma\zeta\sqrt{4\gamma^4 + 48\gamma^3\zeta + 44\gamma^2\zeta^2 + 12\gamma\zeta^3 + \zeta^4}}{2\gamma^2\zeta + \gamma\zeta^2} \right]$$
(107)

To evaluate the payoff to the orange strategy, we wish to evaluate the expected profit of the A, B arbitrageur in equilibrium. We can calculate the equilibrium value of K_{AB} by plugging

the expressions for b_C from (106) into (104). Moreover, from (103), we have:

$$\eta_{AB} = \left(\frac{\gamma}{1+b_C\gamma}\right)(x_A - x_B)$$

Hence,

$$Var\left[\eta_{AB}\right] = 2\left(\frac{\gamma}{1+b_C\gamma}\right)^2 \sigma^2$$

Plugging in b_C from (106), we can then use expression (71) of Proposition 3 to calculate the A, B arbitrageur's expected profits. As with the case in Appendix D.3, the expressions are too convoluted to usefully display, so we keep them in a companion Mathematica notebook. Numerically, we found that, as in the case without price impact, the profits of the "orange" strategy of connecting two connected are always less than the "green" strategy in Appendix D.3 of connecting an isolated peripheral node to a connected peripheral node.

D.5 Star Network Equilibrium

As in the proof of Proposition 1, we have shown that the red strategy is dominated by the blue strategy, and the orange strategy is dominated by the green strategy. Thus, a star network equilibrium exists whenever the blue strategy pays higher than the green strategy; that is, conditional on N-1 arbitrageurs forming a star, the final arbitrageur finds it optimal to complete the star. Since we showed that the red strategy is dominated by the blue strategy, we need only to show that the payoff to the A, B arbitrageur in the green strategy, characterized in Subsection D.3, is lower than (75), the payoff to an arbitrageur connecting a single isolated node to the center.

While the outcomes are complex, expected profits in both cases are homogeneous of degree 1 in the pair (ζ, γ) ; that is, if we scale up both ζ and γ by a constant, all profits scale up by the same factor. Thus, as in the main text, the sole determinant of whether the blue strategy's payoff dominates the green strategy's payoff, as N becomes large, is the ratio $\frac{\zeta}{\gamma}$. Using Mathematica, we verified that the ratio of profits between the two strategies is a simple monotone decreasing function of $\frac{\zeta}{\gamma}$, and the blue strategy has higher payoff whenever:

$$\frac{\zeta}{\gamma} < \frac{1}{3} \left(-4 + \sqrt[3]{80 - 30\sqrt{6}} + \sqrt[3]{10 \left(8 + 3\sqrt{6} \right)} \right)$$

Where the RHS is approximately equal to 1.07395. This proves Proposition 2.

D.6 Proof of Proposition 3

In general, as in (4) in the main text, arbitrageurs' profit is:

$$V(z_{ij}) = z_{ij}\Delta p_{ji} - \frac{\zeta}{2} (z_{ij})^2$$
(108)

Consider Δp_{ji} as a function of z_{ij} , and differentiate with respect to z_{ij} ; a necessary condition for optimality is:

$$z_{ij}\frac{\partial\Delta p_{ji}}{\partial z_{ij}} + \Delta p_{ji} - \zeta z_{ij} = 0$$

From (69) we have that $\frac{\partial \Delta p_{ji}}{\partial z_{ij}} = -K$. Thus, rearranging, for any pair $(z_{ij}, \Delta p_{ji})$ to satisfy arbitrageurs' first-order condition, it must satisfy:

$$-Kz_{ij} - \zeta z_{ij} + \Delta p_{ji} = 0$$

Solving for z_{ij} , we get (70).

To calculate profits, we then express z_{ij} and Δp_{ji} as functions of η_{ij} . We have:

$$z_{i,j} = \frac{\Delta p_{j,i}}{K+\zeta} = \frac{-Kz_{i,j} + \eta_{i,j}}{K+\zeta}$$
$$\implies z_{ij} = \frac{\eta_{ij}}{2K+\zeta}$$

Using (70), we then have:

$$\Delta p_{ji} = \frac{K + \zeta}{2K + \zeta} \eta_{ij}$$

Thus, we can plug in to (108) to evaluate arbitrageur profits as a function of η_{ij} :

$$V(\eta_{ij}) = \left(\frac{1}{2K+\zeta}\eta_{i,j}\right) \left(\frac{K+\zeta}{2K+\zeta}\eta_{i,j}\right) - \frac{\zeta}{2} \left(\frac{1}{2K+\zeta}\eta_{i,j}\right)^2$$

This simplifies to:

$$=\frac{1}{4K+2\zeta}\eta_{ij}^2$$

Taking expectations over η_{ij} , and using that $E[\eta_{ij}] = 0$, we have:

$$E\left[V\left(\eta_{ij}\right)\right] = \frac{E\left[\eta_{ij}^{2}\right]}{4K + 2\zeta} = \frac{Var\left[\eta_{ij}\right]}{4K + 2\zeta}$$

E Entry

In the main text, we assumed there are exactly N - 1 arbitrageurs. In this appendix, we construct an entry game in which, for some entry costs, exactly N - 1 arbitrageurs choose to enter in equilibrium.

We assume there is an infinite number of potential arbitrageurs a = 1, 2, ... Arbitrageurs sequentially decide whether to pay entry cost C and enter the market. Once one arbitrageur decides not to enter, the entry phase concludes, and arbitrageurs play the network game in the main text, simultaneously selecting a pair of exchanges to link. An equilibrium in this game specifies a set of beliefs under which arbitrageurs' entry decisions maximize their expected utility.

Our main result is that there exist entry costs and beliefs that sustain equilibria in which exactly N - 1 arbitrageurs enter. In particular, for sufficiently small values of ϵ , suppose the entry cost C satisfies:

$$C < \frac{\gamma^2}{2\zeta} \left(\frac{\zeta}{\gamma+\zeta}\right)^2 \sigma^2 - \epsilon \tag{109}$$

that is, C is lower than the expected arbitrageur payoff in the "blue" strategy characterized in Appendix C.3.1 by some small ϵ . Then we can construct beliefs which sustain an equilibrium in which exactly N - 1 arbitrageurs enter.

The beliefs we will consider are simple. Arbitrageurs $1 \dots N - 1$ believe that exactly N - 1 arbitrageurs will enter, and they will form a star network; under these beliefs, it is rational for all N - 1 arbitrageurs to enter, since their payoff (for sufficiently large N) is greater than the cost of entry, by the construction of C in (109). The Nth arbitrageur believes that, if she enters, the other N - 1 arbitrageurs will still form a star network. This implies that the Nth arbitrageur has only two possible strategies: she can compete with an existing arbitrageurs, playing the "red" strategy characterized in Appendix C.3.2; or she can play the "orange" strategy, linking two peripheral nodes. The key is then to show that, for small ϵ , neither strategy available to the Nth arbitrageur has expected payoff higher than the entry cost C.

This is true, since we showed in the proof of Proposition 1 in Appendix C that the payoffs of the "red" and "orange" are both lower than the payoff of the blue strategy in the large-Nlimit. Thus, there exists some ϵ such that the blue strategy has positive payoffs net of entry cost, but no other strategies do, in the large-N limit. Under this choice of ϵ , and the belief that N-1 arbitrageurs will form a star network, the Nth arbitrageur expects negative profits net of entry cost from entry, and thus chooses not to enter.

Intuitively, what is driving this equilibrium construction is that a star network with N

exchanges has N - 1 "spokes", implying that there is a discrete decrease in expected payoffs when going from N - 1 to N arbitrageurs. It is thus possible to pick an entry cost where N - 1 arbitrageurs are willing to enter, but not an Nth. Analogously, for lower entry costs, we can construct equilibria in which for example 2(N - 1) arbitrageurs enter, by setting the entry cost such that the "red" strategy of two arbitrageurs has positive expected payoffs, but further crowding is unprofitable; equilibria involving various multiples of (N - 1) are thus possible.

Note also that the network formation game may have non-star-shaped equilibria; our equilibrium construction in the entry game relies on the assumption that arbitrageurs believe the star network equilibrium will be played. In other words, we have constructed an entry game equilibrium which supports the entry of exactly N - 1 arbitrageurs, but there may be other entry equilibria supported by other beliefs about what arbitrageurs will do after entering.

F Data Cleaning

F.1 Cryptotick Data

This section introduces our Cryptotick data cleaning process. The raw dataset contains hourly data for each trading pair on each exchange, including price and volume variables. Our main goal is to create a daily dataset at the coin-exchange pair level, incorporating price and volume variables for each coin on each exchange for each day. We follow five steps:

- 1. Aggregating data at the daily level. For each trading pair on each exchange within a given day, we aggregate the hourly data by computing the volume-weighted average price of each open hour and summing the total trade volume across all hours.
- 2. Focusing on top coins and common trading pair. We restrict our sample in two ways: by the cryptocurrency in the trading pair, and by the denominator that the cryptocurrency is traded against. Since many coins are not actively traded, we first limit our sample to trading pairs that involve the top 200 cryptocurrencies ranked by coinmarketcap.com on June 1 in any year from 2017 to 2023. We further restrict our sample to three types of trading pairs: pairs involving one of 27 major fiat currencies;²⁰ pairs involving BTC or ETH, the two largest cryptocurrencies by market capitalization; and pairs involving one of the three major stablecoins (USDT, USDC, and BUSD).
- 3. Converting prices and volumes to USD terms. For fiat-denominated trading pairs, we convert prices using same-day USD-fiat exchange rates. For crypto-denominated pairs, we convert prices using daily cryptocurrency and stablecoin prices from Yahoo Finance.
- 4. Aggregating data at the coin level. For each of the top cryptocurrencies on each exchange and each day, we aggregate data across all trading pairs. We compute the volume-weighted average price across all relevant trading pairs and sum the total trade volume. We exclude stablecoins and fiat currencies from the final dataset.
- 5. Winsorizing and imputating data. We winsorize prices at 2 times the median price on the upper bound and 0.5 times the median price on the lower bound for each coin on each day. This ensures proper measurement of dispersion and prevents extreme values from inflating standard deviations of log prices. Additionally, we impute missing

 $^{^{20}{\}rm These}$ 27 major fiat currencies are: NZD USD KRW JPY CNY IDR SGD VND TWD AUD PKR ZAR TRY MXN BRL CHF ILS PLN GBP RUB EUR CAD HKD INR SAR AED SEK.

data for coins that are listed on an exchange but have no recorded trading activity for certain days.²¹ We assign missing prices as "NA" and set trade volumes to zero.

Finally, we obtain daily prices and volumes in USD terms for each coin listed on exchanges. The coin-exchange level dataset from step 5 serves as the primary dataset for most of our analyses. For some analyses, we further aggregate data to the coin level or exchange level by computing the volume-weighted average price and total trade volume.

F.2 Cross-Exchange Flow Data

F.2.1 Gathering Exchange Wallet Addresses

We first manually collect all available exchange wallet addresses from Etherscan, identifying addresses affiliated with specific exchanges. Specifically, we search for the exchange's name on Etherscan, which returns all addresses with labels associated with that exchange. The labels affiliated with exchanges mostly have the following formats: Exchange #index, Exchange: Hot Wallet #index, Exchange: Cold Wallet #index, Exchange: Deposit Funder #index, Exchange: Deployer #index, and Exchange: Withdrawals #index, where #index is a numeric value. If the labels do not match any of these categories, we use our discretion and collect only addresses that are highly likely to belong to exchanges. For example, some exchange addresses that have been hacked are labeled as Exchange Hacked #index, and we include these addresses as well.

F.2.2 Collecting Cross-Exchange Token Flows

As we discussed in the main text, sending tokens between exchanges generally involves a trader-specific "deposit" address, controlled by exchange B, which is later "swept" into B's main wallet. The process of an arbitrage can be depicted as follows:

Exchange A wallet address $\xrightarrow{\operatorname{txn} X}$ Trader's deposit address on Exchange $B \xrightarrow{\operatorname{txn} Y}$ Exchange B wallet address (110)

The first transaction in (110), "txn X", cannot be immediately identified as an arbitrage transaction because deposit addresses are not labelled on etherscan, and thus we cannot distinguish them from private wallets that do not belong to exchanges. However, we can

²¹Specifically, if a coin appears on an exchange but has missing data for some days between its first and last observed trading date, we impute these observations. As a result, the number of observations increased by 19%, from 9,747,409 to 12,034,327.

infer that the recipient address was likely an exchange deposit address – and thus transaction X was likely an arbitrage transaction – if we later observe the address sending the received tokens into exchange B's labelled wallet address, which we label as "txn Y" in (110).

We first gather all send and receive transactions from all identified exchange wallets, using the Etherscan API. We then identify pairs of such transactions that are likely to be arbitrage transactions as follows.

- 1. We first find all withdrawal transactions for all identified exchange wallets that is, transactions in which tokens are sent out of the exchange wallet address to any other address.
- 2. For each withdrawal with receipient address Z, we attempt to find a deposit transaction in which Z sends the same coin to another exchange's labelled wallet address.

We additionally impose the restrictions that transaction amounts range from 99% to 101%, and the deposit transactions occur no later than 48 hours after the withdrawal transactions. By following this procedure, we identify all withdrawals where the recipient later initiates a deposit transaction to another exchange wallet. As a result, approximately 8% of deposits and withdrawals are classified as arbitrage trades.

G Calculation of Network Centrality Measures

The section formally defines our three centrality definitions used in Table 2 – degree, betweenness, and eigenvector centrality. As we discussed in Subsection 5.2, we construct an unweighted network among exchanges, defining two exchanges as being connected if bilateral arbitrage flows between two exchanges are large relative to the total arbitrage flows in and out of each exchange: formally, if $ArbVol_{e,\tilde{e}}$ is at least 10% of the smaller of $ArbVol_e$ and $ArbVol_{\tilde{e}}$. We can then represent the resultant network of exchanges using the *adjacency matrix*, which is an $N \times N$ matrix where N denotes the number of exchanges. The element M_{ij} is equal to 1 if exchanges are linked, and 0 otherwise. Diagonal elements of M_{ij} are all set to 0.

The centrality measures we use are then defined as follows. The *degree centrality* of an exchange i is defined simply as the number of edges the exchange has with other exchanges. Formally, the degree centrality of exchange i is:

$$C_i^{Deg} = \sum_{j \neq i} \boldsymbol{M}_{ij}$$

As is standard in the literature, we normalize C_i^{Deg} to sum to one across exchanges.

The betweenness centrality of an exchange counts the fraction of shortest paths between two randomly chosen exchanges that pass through a given exchange. Let s_{jk} denote the number of shortest paths between exchange j and k, and let $s_{jk}(i)$ denote the number of these shortest paths that go through exchange i. The betweenness centrality of exchange i is then:

$$C_{i}^{Bet} = \sum_{\substack{j,k,\\j \neq k \neq i}} \frac{s_{jk}\left(i\right)}{s_{jk}}$$

Again, we normalize C_i^{Bet} to sum to one across exchanges.

The eigenvector centrality of an exchange is defined based on finding the exchange's component of the eigenvector of the adjacency matrix M with the largest eigenvalue. In contrast to degree centrality, which simply counts edges, eigenvector centrality also depends on the quality of a node's connections (Newman, 2008). Again, we normalize the eigenvector values to sum to one across exchanges.

H Robustness Checks

H.1 Delistings

In the main text, we conduct difference-in-differences tests using listing events. This appendix analyzes *delistings*, where central exchanges stop trading a coin. Delistings are significantly rarer than listings: during our sample period, we only observe 69 and 17 delistings on Binance and Coinbase, respectively. Our model implies that delistings should have the opposite effects of listings on market outcomes: after delistings, price dispersion should increase; arbitrage volume and trade volume should decrease; and peripheral exchanges should be less likely to list coins for trade, or may follow central exchanges' delisting decisions.

Columns (1) and (2) in Appendix Table A.2 show the effects of delistings on price dispersion. To avoid compositional effects from exchanges entering or exiting the market on measured price dispersion, we use a balanced panel for coins delisted by central exchanges. Specifically, we keep only exchanges that remain active for at least 30 days after a central exchange delisting. For never-treated coins, we include all exchanges. Results remain similar when all peripheral exchanges are included in the sample. Price dispersion increases significantly following Binance delistings, while the effect is less pronounced for Coinbase delistings.

Columns (3) to (6) show the arbitrage volume results. Similarly, arbitrage volume mostly decreases on both intensive and extensive margins significantly following both Binance and Coinbase delistings.

Columns (7) to (8) show the listing following results. We find that coin delistings from Binance lead to a statistically significant decrease in the number of peripheral exchanges listing the coin, whereas the effect is less significant for Coinbase delistings.

Appendix Table A.3 shows the effects of delistings on trade volumes. We include all peripheral exchanges, regardless of whether they delist the coin; however, results remain similar when restricting the sample to exchanges that remain active for at least 30 days after a central exchange delisting. Due to the limited number of delistings, estimating clustered standard errors is infeasible, so we use robust standard errors instead. We find that Binance delistings lead to statistically significant decreases in trade volumes on peripheral exchanges. For Coinbase, the point estimates generally indicate a decline in volumes, though the effects are not statistically significant.

H.2 "Logs with Zeros"

When we test the effects of central exchange listings on arbitrage and trade volume in the main text, we separately estimate the extensive and intensive margin effects. Here we show the robustness of our results to two different ways of handling zeros.

Log-plus-*a* **Transformation.** First, we test the robustness of our results using the commonly applied $\log (a + x)$ transformation. We do not do this in the main text since Chen and Roth (2024) recommend against using this transformation, as it combines intensive and extensive margins in unintuitive ways. However, given that we find statistically significant effects on both margins, our results should remain robust to using $\log (a + x)$. This is confirmed in Appendix Tables A.4, A.5, and A.6, where we repeat the specifications of Tables 3, 4, and 5 for various choices of a. In all cases in Appendix Tables A.4 and A.5, as well as in the majority of cases in Appendix Table A.6, our effects remain significant with unchanged signs.

Balanced Panel. Another approach to handling zeros is to run regressions using balanced panels, keeping only coin-exchange pairs that have positive arbitrage or trade volume each day within a 60-day window around the central exchange's listing date. This subset of the data naturally forms a balanced panel. In contrast, our baseline specification estimates intensive margin effects using an unbalanced panel of coin-exchange pairs, where zero-volume observations around treatment dates are dropped. Appendix Table A.7 shows the results. Columns (1) and (2) report the arbitrage volume results, Columns (3) and (4) report the trade volume results, and Columns (5) and (6) report the trade volume results with exchange correlation heterogeneity. The results in Columns (1) to (4) remain significant with unchanged sign. The correlation tests in Columns (5) and (6) are mostly statistically insignificant, though they have signs consistent with our baseline estimates.

H.3 Attention

An important alternative explanation for our results is that, instead of the arbitrage hypothesis we propose and test, investors' *attention* to coins is a confounding variable which may influence (or be influenced by) listings and trade volumes. In this subsection, we measure investor attention, and evaluate two ways in which attention could confound our results.

Following studies such as Da, Engelberg and Gao (2011), we collect the daily Google search volumes for the ticker symbols of coins in our sample. This data provides an index ranging from 0 to 100, based on the search volumes for the keywords during the query period. When the query period is too long, Google only provides weekly or monthly data. Therefore,

we follow the literature and obtain a consistent daily Google search volume index by stitching each semiannual's daily data (Hoopes, Reck and Slemrod, 2015; Li et al., 2023). The detailed data collection process is described in Appendix H.3.3 below.

We use the log of 1 plus the search volume index to create a coin-day level measure of attention, which we refer to as $Attention_{c,t}$.²² Note that $Attention_{c,t}$ is a relative measure, indexed at the coin level, so cannot be compared across coins; we will thus use coin fixed effects in all specifications.

We first conduct a simple DID regression to illustrate how treated coin's attention changes around central exchange listings:

$$Attention_{c,t} = \beta_1 PreListing(33-93 \text{ days})_{c,t} + \beta_2 PreListing(3-33 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} + \beta_4 Listing(0-30 \text{ days})_{c,t} + \beta_5 Listing(> 30 \text{ days})_{c,t} + \delta_c + \eta_t + \epsilon_{c,t}$$

$$(111)$$

The results are shown in Appendix Table A.8. There are two main findings. First, the coefficient on $PreListing(3-33 \text{ days})_{c,t}$ is positive and significant for Coinbase: coin level attention tends to increase in the run-up to central exchange listings. This is intuitive: central exchanges are likely to list coins that are increasing in popularity. Second, the coefficient on $Listing(0-30 \text{ days})_{c,t}$ is also positive and significant. This is consistent with the idea that central exchange listings may increase attention to treated coins.

We now discuss two ways in which investor attention could influence the interpretation of our results.

H.3.1 Attention Drives Everything

First, we test the strong hypothesis that investor attention may be the core variable driving all outcomes we observe: variation in investor attention to coins may drive central exchange listings, peripheral exchange listings, and trade volumes. We call this the "attention drives everything" hypothesis. This hypothesis would be problematic for our core results, since it would imply that the effects we attribute to central exchange listings are in fact not causal: rather, they simply reflect the effects of increased investor attention driving changes in all outcome variables.

This hypothesis is comparatively easy to provide evidence against. Appendix Figures A.2 to A.7 show that the relationship between central exchange listings and outcomes is very sharp; this is difficult to reconcile with a hypothesis in which gradual increases in attention

 $^{^{22}}$ The existing literature uses either the level of log of 1 plus the search volume, or the abnormal change in log of 1 plus the search volume as a measure of attention. We use log of 1 plus the search volume since it is naturally analogous to log trade volume, which is the dependent variable in our main specifications.

smoothly increase central exchange listing probabilities, arbitrage volume, trade volumes, and peripheral exchange listing probabilities, but central exchange listings do not directly impact either outcome.

However, it could be possible that pre-listing attention partially accounts for the listing effect that we have found. Attention may influence central exchanges' listing decisions: that is, central exchanges may have a higher propensity to list coins which have rising investor attention. This is both intuitively plausible, and consistent with evidence in Appendix Table A.8. To alleviate this concern, we perform a "matched controls" test.

For each "treated" coin listed by a central exchange, we find a control coin, that is not listed by a central exchange, but has similar level and time series patterns in attention to the treated coin over the quarter prior to the central exchange's listing. Since treated and control coins do not differ in attention prior to listing, under the "attention drives everything" hypothesis, we would not expect treated and control coins to differ substantially in outcomes after central exchange listings.

We search for matched coins based on two criteria: average log of 1 plus trade volume; and correlations in moving average of attention. The precise matching criteria and process are described in Appendix H.3.4. In the end, we construct balanced coin-day and coinexchange-day panels for Binance and Coinbase listings, respectively. For each treated coin (coin-exchange pair), we have a control coin (coin-exchange pair) with similar pre-listing attention patterns. Treated and control groups are comparable before listings. The matching rate is around 60% at coin-exchange level and 90% at coin level.

We use the balanced panel and run the same regressions as in Specifications (17) and (18), further adding matched-pair×day fixed effects to compare treated with control coins (coin-exchange pairs) in each matched pair. For (17), we use the coin-day panel, and for (18), we use the coin-exchange-day panel. These matched pairs of coins (coin-exchange pairs) share similar ex-ante attention characteristics, with the difference being that one experiences a central exchange listing while the other does not.

Appendix Tables A.9 and A.10 show the results. The results show that, when two coins (coin-exchange pairs) A and B have very similar characteristics in pre-listing attention, but coin A is listed while coin B is not, coin A experiences a relative dispersion decrease, arbitrage volume increase, number of exchanges listing the coin increase, and trade volume increase. This would not be expected under the hypothesis that attention is the primary driver of volume, and that listings have no direct effect on volumes after controlling for prelisting attention.

H.3.2 The "Attention Channel"

Relative to the strong hypothesis that "attention drives everything", a weaker hypothesis is that central exchange listings may causally affect outcomes, but attention may serve as a *channel* through which central exchange listings influence outcomes, distinct from our arbitrage channel. That is, when Binance or Coinbase lists a coin, market participants pay increased attention to the coin, leading them to trade these coins more actively on peripheral exchanges, also giving other peripheral exchanges greater incentives to list these coins. This is consistent with the finding in Appendix Table A.8, that attention increases after central exchange listings. We cannot fully rule out the "attention channel" hypothesis. We will show evidence suggesting that the attention channel may exist, but it has difficulty fully explaining the magnitudes of the effects we observe.

First, notice that this would not explain all of our stylized facts: it does not explain why price dispersion falls after centralized exchange listings, why the volume increase and listing following effects are larger for peripheral exchanges with stronger price correlations with central exchanges, and why cross-exchange flows on the blockchain increase following coin listings.

Second, we can directly control for the attention channel by adding $Attention_{c,t}$ in the regression. Notice in this subsection, we still use the balanced matching panel and add matched-pair×day FE to compare coin (coin-exchange pair) with similar pre-listing attention.

If the entirety of the effect of central exchange listings were explained by the induced postlisting increase in coin attention, we would expect that controlling for the mediating variable of attention would cause our DID estimates to become insignificant. In Appendix Tables A.11 and A.12, we add attention as a control to the attention-matched sample regressions of Appendix Tables A.9 and A.10; results do not change much, suggesting that the "attention channel" quantitatively does not seem to explain the entirety of our results.

What fraction of the results we observe might be attributable to the attention channel? We do a simple back-of-envelope exercise to approximately quantify this effect on trade volume. We first estimate the effect of attention on trade volumes using a simple panel regression, as shown in Appendix Table A.13. We then estimate the effects of central exchange listings on coin level attention, as shown in Appendix Figure A.8, with attention as the dependent variable. Under the (admittedly strong) assumption that the panel regression coefficient captures the causal effect of coin attention increases on trade volume, we can then quantify the magnitude of the attention channel, by multiplying our estimated effects of central exchange listings on attention, by our estimates of the effect of attention on trade volumes and other outcomes. In Appendix Figure A.9, the blue line plots the results of this back-of-envelope exercise. We find that the attention channel accounts for only about 15% of the total volume effect (red line). Intuitively, central exchange listings increase attention by around 41% and 33% at the peak, and we estimate that a 1% increase in attention increases trade volume by around 0.44% and 0.59%, for Binance and Coinbase listings, respectively. The results are similar if we further allow these effects vary by coin-exchange pairs, as shown in the green line.²³

To summarize, we think that attention is likely to be one of the channels through which listings increase trade volumes: listings have a statistically and economically significant effect on coin attention, and the idea that attention causally impacts trade volumes is both intuitively plausible, and consistent with the panel-regression attention coefficient in Appendix Table A.13. However, based on our control test and back-of-envelope accounting exercise, we think it is unlikely that the attention channel quantitatively explains the majority of the effect of listings on trade volumes.

H.3.3 Attention Measurement

In this section, we describe the process of collecting the daily Google Search Volume Index (SVI). The SVI is defined as:

$$SVI_{i,t,W} = \left(\frac{Keyword_Search_{i,t}/Global_Search_t}{\max_{t \in W} \{Keyword_Search_{i,t}/Global_Search_t\}}\right) \times 100$$

where i, t, and W denote keyword, time, and query window, respectively. The numerator, $Keyword_Search_{i,t}/Global_Search_t$, is the keyword's search volume normalized by the global total search volume at the corresponding time. This measure has been widely used in the existing literature to quantify attention. The denominator,

$$\max_{t \in W} \{Keyword_Search_{i,t}/Global_Search_t\}$$

is the keyword's maximum relative search volume during the query window. Consequently, the index ranges from 0 to 100.

One challenge is that Google does not provide daily SVI when the query window is too long. Moreover, we cannot directly use daily SVI from consecutive shorter query windows because the denominator in each period differs. To address this, we follow the literature and

 $^{^{23}}$ Specifically, we estimate the volume-attention coefficient for each matched pair and then estimate the listing effects on attention using the treated and control coin-exchange pairs within that matched pair. The attention channel for each matched pair is quantified by multiplying the estimated effects of central exchange listings on attention with the estimates of the effect of attention on trade volumes. The heterogeneous attention channel is then calculated by averaging these effects across all matched pairs.

apply a stitching approach to recover the daily numerator values (Hoopes, Reck and Slemrod, 2015; Li et al., 2023). We then normalize these values in the same way as Google, treating the maximum as 100, to construct a consistent daily SVI for the entire period.

Specifically, we recursively collect the daily SVI for each two-quarter window W_j , separated by a one-quarter gap (e.g., 2017 Q1–Q2, 2017 Q2–Q3, etc.). Each pair of consecutive periods has one quarter of overlapping data. Using this overlap, we recover the ratio of the denominators $\frac{\max_{t \in W_j} \{Keyword_Search_{i,t}/Global_Search_t\}}{\max_{t \in W_{j-1}} \{Keyword_Search_{i,t}/Global_Search_t\}}$ by averaging the daily SVI ratios in the overlapping quarter:

 $\frac{\max_{t \in W_j} \left\{ Keyword_Search_{i,t}/Global_Search_t \right\}}{\max_{t \in W_{j-1}} \left\{ Keyword_Search_{i,t}/Global_Search_t \right\}} = \frac{1}{\# overlapping_days} \sum_{t=1}^{\# overlapping_days} \frac{SVI_{i,t,W_{j-1}}}{SVI_{i,t,W_j}}$

Using this ratio, we recover numerator values for W_j by multiplying SVI_{i,t,W_j} in the nonoverlapping quarter by the above ratio:

$$\begin{split} & Keyword_Search_{t} \\ &= SVI_{i,t,W_{j}} \times \frac{\max_{t \in W_{j}} \{Keyword_Search_{i,t}/Global_Search_{t}\}}{\max_{t \in W_{j-1}} \{Keyword_Search_{i,t}/Global_Search_{t}\}} t \in W_{j} \text{ and } t \notin W_{j-1} \end{split}$$

By recursively applying this procedure, we obtain the daily $Keyword_Search_{i,t}/Global_Search_t$ for the entire period, which is $Keyword_Search_{i,t}/Global_Search_t$ divided by the maximum index in the first quarter. Finally, to derive a daily SVI, $SVI_{c,t}$, we normalize $Keyword_Search_{i,t}/Global_Search_t$ by treating the maximum value as 100.

Finally, our attention measure, $Attention_{c,t}$, is defined as the log of 1 plus the search volume index:

$$Attention_{c,t} = \log (1 + SVI_{c,t})$$

H.3.4 Matched Controls Construction

This appendix describes how we construct our attention-matched sample of coins listed by central exchanges. For each "treated" coin listed by a central exchange, we attempt to find a "matched control" coin with similar attention patterns to the treated coin prior to central exchange listing; we then compare treated and control coins in a variety of specifications.

Ideally, we attempt to find control coins which match our treated coins in terms of levels, and time-series patterns in attention. Our attention measure is indexed at the coin level, so it cannot be meaningfully compared across coins. Instead, we match control coins based on average log of one plus trade volume, over the one quarter prior to the treated coin's listing. To mitigate announcement effects, we use data from the period between T - 93 and T - 3,

where T represents the central exchange listing date. We expect that two coins with similar average trade volume should have similar attention level.

The second criterion is the correlation between the one-week moving average of $Attention_{c,t}$ in the quarter (T - 93 to T - 3) prior to the central exchange listing for the treated coin and the control coin. This attention correlation aims to capture low-frequency trends in attention prior to listing. For example, if a treated coin shows an upward attention trend prior to listing, we try to match it with a control coin exhibiting a similar upward trend in the same time window.

We use these two criteria to construct a balanced panel dataset at both the coin-day and coin-exchange-day levels. For the coin-day level balanced panel, treated and control coins must have a trade volume difference of less than 2, and an attention correlation greater than 0.5. If multiple candidates meet these criteria, we select the control coin with the highest attention correlation. For the coin-exchange-day level balanced panel, treated and control coin-exchange pairs must be on the same exchange and also meet the preceding conditions.

The matching rate and the full summary statistics are shown in Appendix Table A.14. The matching rate is around 60% at the coin-exchange level, and 90% at coin level. Treated and control coins (coin-exchange pairs) are quite comparable. Another way to test our matching procedure is to run the same DID specification for $Attention_{c,t}$ using our balanced sample. Appendix Figure A.8 shows the results, and there are no pre-trends: attention in the control and treatment groups appears to trend identically prior to listing, suggesting our attention-matched control group is constructed fairly well.

I Appendix Figures and Tables

Figure A.1: Arbitrage Flow Network

The figure shows the arbitrage volume network of crypto exchanges. We assume two exchanges e and \tilde{e} are "connected" whenever the arbitrage flows between e and \tilde{e} are a large fraction of total arbitrage flows involving either exchange: formally, if $ArbVol_{e,\tilde{e}}$, is at least 10% of the smaller of $ArbVol_e$ and $ArbVol_{\tilde{e}}$. The size of each node is proportional to the number of edges connected to the node. Data sources: Cryptotick, Etherscan.



Figure A.2: Central Exchange Listings and Price Dispersion: Event-study Estimation

The figure shows estimates from Specification (19):

$$Dispersion_{c,t} = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$

along with 95% confidence intervals. The dependent variable is coin-day level price dispersion, calculated as the standard deviation of log prices across exchanges. For treated coins, we only use incumbent exchanges that list coin c at least 30 days before its listing on a central exchange. For never-treated coins, we use all exchanges. δ_c represents coin fixed effects, and η_t represents day fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin and time level. Data source: Cryptotick.



Figure A.3: Central Exchange Listings and Arbitrage Volume Dummy: Event-study Estimation

The figure shows estimates from Specification (19):

$$I(ArbVol_{c,t} > 0) = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$

along with 95% confidence intervals. The dependent variable is a coin-day level dummy for arbitrage volume being greater than 0. For treated coins, we consider arbitrage flows that involve the central exchange, and incumbent exchanges which list the coin at least 30 days before the central exchange lists. For never-treated coins, we add all arbitrage flows. δ_c represents coin fixed effects, and η_t represents day fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin and time level. Data source: Cryptotick, Etherscan.



Figure A.4: Central Exchange Listings and Log Arbitrage Volume: Event-study Estimation

The figure shows estimates from Specification (19):

$$\log (ArbVol_{c,t}) = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$

along with 95% confidence intervals. The dependent variable is coin-day level log arbitrage volume. For treated coins, we consider arbitrage flows that involve the central exchange, and incumbent exchanges which list the coin at least 30 days before the central exchange lists. For never-treated coins, we add all arbitrage flows. δ_c represents coin fixed effects, and η_t represents day fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin and time level. Data source: Cryptotick, Etherscan.



Figure A.5: Central Exchange Listings and Trade Volume Dummy: Event-study Estimation

The figure shows estimates from Specification (20):

$$I(TradeVol_{c,e,t} > 0) = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$

along with 95% confidence intervals. The dependent variable is a coin-exchange-day level dummy for trade volume being greater than 0. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. $\delta_{c,e}$ represents coin-exchange fixed effects, η_t represents day fixed effects, and $\gamma_{e,t}$ represents exchange-time fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin-exchange pair and time level. Data source: Cryptotick.



Figure A.6: Central Exchange Listings and Log Trade Volume: Event-study Estimation

The figure shows estimates from Specification (20):

$$\log(TradeVol_{c,e,t}) = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$

along with 95% confidence intervals. The dependent variable is coin-exchange-day level log trade volume. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. $\delta_{c,e}$ represents coin-exchange fixed effects, η_t represents day fixed effects, and $\gamma_{e,t}$ represents exchange-time fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin-exchange pair and time level. Data source: Cryptotick.



Figure A.7: Central Exchange Listings and Peripheral Exchange Listings: Event-study Estimation

The figure shows estimates from Specification (19):

$$\Delta \# Exchanges_{c,t} = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_c + \eta_t + \epsilon_{c,t}$$

along with 95% confidence intervals. The dependent variable is the number of exchanges that list coin c between day t-1 and t. We filter the sample to include only coins that were listed on any exchange for at least 30 days before being listed on central exchanges or were not listed by central exchanges, eliminating the mechanical effects associated with coins initially listed on central exchanges. δ_c represents coin fixed effects, and η_t represents day fixed effects. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin and time level. Data source: Cryptotick.



Figure A.8: Central Exchange Listings and Coin Attention: Event-study Estimation (Matched Sample)

This figure shows estimates from the following specification:

$$Attention_{c,t} = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_c + \eta_t + \kappa_{matchid,t} + \epsilon_{c,e,t}$$

along with 95% confidence intervals. The dependent variable is the log of 1 plus the Google search volume index. δ_c represents coin fixed effects, η_t represents day fixed effects, and $\kappa_{matchid,t}$ represents matched pair×time fixed effects. Each treated coin is matched with a control coin based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Observations exactly 30 days before central exchange listings are set as the reference group. Standard errors are clustered at the coin and time level. Data source: Cryptotick, Google Search.



Figure A.9: Central Exchange Listings, Trade Volumes, and Attention: Back-of-Envelope Quantification

This figure shows the listing effects on trade volume by highlighting the attention channel and comparing the effects with the aggregate listing effects. The dependent variable is coin-exchange-day level log trade volume. The red line represents the aggregate effects estimated from Specification (20):

$$log(Volume_{c,e,t}) = \sum_{k=-31}^{31} \beta_k \times treat_{c,k,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \kappa_{matchid,t} + \epsilon_{c,e,t}$$

The blue line shows the effects driven by the attention channel, calculated by multiplying the volume-attention relationship with the dynamic attention effect. The green line further relaxes by estimating the effect for each coin-exchange pair individually and then show the average effects across all pairs. $\delta_{c,e}$ represents coin-exchange fixed effects, η_t represents day fixed effects, $\gamma_{e,t}$ represents exchange-time fixed effects, and $\kappa_{matchid,t}$ represents matched pair-time fixed effects. Each treated coin-exchange pair is matched with a control pair based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Data source: Cryptotick, Google Search.



Table A.1: Price Gap Reversion

This table shows estimates of Specification (14), (15), and (16) for all coin-exchange pairs:

$$\Delta PriceGap_{c,e,t} = \beta^{PriceGap} PriceGap_{c,e,t-1} + \epsilon^{PriceGap}_{c,e,t}$$
$$\Delta p_{c,e,t} = \beta^{p} PriceGap_{c,e,t-1} + \epsilon^{p}_{c,e,t}$$
$$\Delta p_{c,t}^{cen} = \beta^{p^{cen}} PriceGap_{c,e,t-1} + \epsilon^{p^{cen}}_{c,e,t}$$

We follow the suggestion of (Pesaran, 2006) and estimate a consistent mean group estimator for $\beta^{PriceGap}$, β^p , and $\beta^{p^{cen}}$. It follows a two-stage estimation. In the first stage, we similarly estimate the individual equation for each coin-exchange pair, but we also include three lagged cross-sectional averages of dependent variables $\sum_{c,e} \Delta PriceGap_{c,e,t-1}$, $\sum_{c,e} \Delta PriceGap_{c,e,t-2}$, and $\sum_{c,e} \Delta PriceGap_{c,e,t-3}$. In the second stage, we use the mean of individual estimators as the estimate for β and the variance of individual estimators as the estimate for $Var(\beta)$. Columns (1) to (2) are results with change in price gap as the dependent variable. Columns (3) to (4) are results with peripheral exchange price change as the dependent variable. Columns (5) to (6) are results with central exchange price change as the dependent variable. Odd (even) number columns are results treating Binance (Coinbase) as the central exchange. Standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	$\Delta PriceGap$		Δ	Δp	Δp^{cen}		
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	
$PriceGap_{c,e,t-1}$	-0.72^{***} (0.02)	-0.74^{***} (0.03)	-0.67^{***} (0.04)	-0.97^{***} (0.19)	0.05 (0.04)	-0.22 (0.2)	
R ² Observations	0.3 6287565	0.68 3522046	$0.13 \\ 6287565$	0.57 3522046	0.07 6287565	0.5 3522046	

This table shows estimates from the following specification:

$$\begin{split} Y_{c,t} = & \beta_1 Delisting(0\text{--}30 \text{ days})_{c,t} + \beta_2 Delisting(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayDelisting_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. $Delisting(0-30 \text{ days})_{c,t}$ and $Delisting(> 30 \text{ days})_{c,t}$ are dummies for coins delisted within the past 30 days or more than 30 days ago. $PreThreedayDelisting_{c,t}$ accounts for the pre-announcement effect. For price dispersion, we exclude newly exited exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan.

Dependent Variables:	Dispersion		I(ArbVol > 0)		Log(ArbVol)		Net Listings	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Delisting (0-30 days)	0.04^{*}	0.05	0.04	-0.24**	-1.7***	-1.3*	-0.09***	-0.05
	(0.02)	(0.06)	(0.06)	(0.10)	(0.45)	(0.70)	(0.02)	(0.03)
Delisting $(> 30 \text{ days})$	0.08^{**}	0.14	-0.42^{***}	-0.33***	-3.9***	-2.0***	-0.0010	-0.006
	(0.03)	(0.12)	(0.05)	(0.13)	(1.1)	(0.36)	(0.002)	(0.004)
Pre Three-day Delisting	-0.003	0.05	0.17^{*}	-0.22**	1.3^{***}	0.10	-0.15***	-0.04
	(0.02)	(0.05)	(0.09)	(0.11)	(0.30)	(0.56)	(0.03)	(0.03)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted \mathbb{R}^2	0.47	0.46	0.42	0.41	0.59	0.57	0.07	0.07
Observations	827,203	$863,\!173$	$272,\!574$	$272,\!574$	$171,\!601$	$171,\!601$	$1,\!352,\!156$	$1,\!352,\!156$

This table shows estimates from the following specification:

$$\begin{split} Y_{c,e,t} = & \beta_1 Delisting(0\text{-}30 \text{ days})_{c,t} + \beta_2 Delisting(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayDelisting_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variables are trade volume dummy and log trade volume. $Delisting(0-30 \text{ days})_{c,t}$ and $Delisting(> 30 \text{ days})_{c,t}$ are dummies for coins delisted within the past 30 days or more than 30 days ago. $PreThreedayDelisting_{c,t}$ accounts for the pre-announcement effect. Odd and even columns are results based on listings on Binance and Coinbase, respectively. We use robust standard errors. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:		I(Trade	Log(TradeVol)					
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Delisting (0-30 days)	-0.04***	-0.02	-0.04***	-0.02*	-0.38***	0.06	-0.36***	0.05
	(0.01)	(0.02)	(0.01)	(0.009)	(0.12)	(0.11)	(0.10)	(0.09)
Delisting $(> 30 \text{ days})$	-0.10***	0.002	-0.07***	-0.007	-1.1***	-0.13	-1.0***	-0.12
	(0.02)	(0.02)	(0.01)	(0.01)	(0.16)	(0.16)	(0.13)	(0.13)
Pre Three-day Delisting	0.03^{***}	-0.01	0.005	-0.02	0.47^{***}	0.32^{*}	0.45^{***}	0.32^{**}
	(0.01)	(0.02)	(0.008)	(0.01)	(0.11)	(0.17)	(0.10)	(0.16)
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exchange FE-Day FE	No	No	Yes	Yes	No	No	Yes	Yes
Adjusted \mathbb{R}^2	0.48	0.48	0.72	0.72	0.79	0.79	0.85	0.85
Observations	$12,\!034,\!327$	$12,\!034,\!327$	$12,\!034,\!327$	$12,\!034,\!327$	9,747,409	9,747,409	9,747,409	9,747,409

Table A.4: Central Exchange Listings and Log-like Arbitrage Volume

This table shows estimates from Specification (17):

$$\begin{split} \log{(a + ArbVol_{c,t})} = & \beta_1 Listing(0\text{--}30 \text{ days})_{c,t} + \beta_2 Listing(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayListing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variable is log of a plus arbitrage volume, where we use a range of parameter a. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For treated coins, we consider arbitrage flows that involve the central exchange, and incumbent exchanges which list the coin at least 30 days before the central exchange lists. For never-treated coins, we add all arbitrage flows. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, ***, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan.

Dependent Variables:	Log(1 +	ArbVol)	Log(10+ArbVol)		Log(100+ArbVol)		Log(1000+ArbVol)		Log(10000+ArbVol)	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Listing (0-30 days)	5.4^{***}	4.0***	4.4***	3.5***	3.4^{***}	2.9***	2.4***	2.2***	1.4^{***}	1.5^{***}
	(0.64)	(0.44)	(0.53)	(0.37)	(0.42)	(0.30)	(0.32)	(0.24)	(0.21)	(0.17)
Listing $(> 30 \text{ days})$	5.1^{***}	3.6^{***}	4.2^{***}	3.0^{***}	3.2^{***}	2.5^{***}	2.2^{***}	1.8^{***}	1.3^{***}	1.2^{***}
	(0.55)	(0.46)	(0.45)	(0.39)	(0.36)	(0.31)	(0.27)	(0.24)	(0.18)	(0.17)
Pre Three-day Listing	1.8^{**}	2.9^{***}	1.4^{**}	2.5^{***}	1.1^{**}	2.0^{***}	0.75^{**}	1.5^{***}	0.41^{**}	0.97^{***}
	(0.69)	(0.43)	(0.56)	(0.36)	(0.44)	(0.29)	(0.31)	(0.23)	(0.20)	(0.17)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.56	0.54	0.58	0.56	0.60	0.58	0.63	0.60	0.64	0.61
Observations	200,044	256,331	200,044	256,331	200,044	256,331	200,044	256,331	200,044	$256,\!331$

Table A.5: Central Exchange Listings and Log-like Trade Volume

This table shows estimates from Specification (18):

$$\begin{split} \log{(a + TradeVol_{c,e,t})} = & \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variable is log of a plus trade volume, where we use a range of parameter a. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	Log(1+7)	TradeVol)	Log(10+TradeVol)		Log(100+TradeVol)		Log(1000+TradeVol)		Log(10000+TradeVol)	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Listing (0-30 days)	0.76***	0.78***	0.70***	0.73***	0.63***	0.68***	0.55^{***}	0.60***	0.45^{***}	0.51^{***}
	(0.08)	(0.05)	(0.07)	(0.05)	(0.06)	(0.04)	(0.05)	(0.04)	(0.04)	(0.03)
Listing $(> 30 \text{ days})$	1.0^{***}	0.92^{***}	0.91^{***}	0.84^{***}	0.78^{***}	0.75^{***}	0.64^{***}	0.65^{***}	0.46^{***}	0.51^{***}
	(0.10)	(0.06)	(0.09)	(0.05)	(0.08)	(0.05)	(0.06)	(0.04)	(0.05)	(0.03)
Pre Three-day Listing	0.59^{***}	1.0^{***}	0.55^{***}	0.94^{***}	0.51^{***}	0.87***	0.45^{***}	0.78***	0.35***	0.66***
	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.04)	(0.05)
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exchange FE-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted \mathbb{R}^2	0.85	0.86	0.86	0.87	0.86	0.86	0.86	0.86	0.85	0.84
Observations	$4,\!895,\!893$	9,004,676	$4,\!895,\!893$	9,004,676	$4,\!895,\!893$	9,004,676	$4,\!895,\!893$	9,004,676	$4,\!895,\!893$	$9,\!004,\!676$

This table shows estimates from Specification (21):

$$\begin{split} \log{(a + TradeVol_{c,e,t})} = & \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(0-30 \text{ days})_{c,t} \times Correlation_e + \\ & \beta_3 Listing(> 30 \text{ days})_{c,t} + \beta_4 Listing(> 30 \text{ days})_{c,t} \times Correlation_e + \\ & \beta_5 PreThreedayListing_{c,t} + \beta_6 PreThreedayListing_{c,t} \times Correlation_e + \\ & \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variable is log of a plus trade volume, where we use a range of parameter a. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. $Correlation_e$ is the return correlation of Bitcoin between the central exchange and the peripheral exchange e using the entire time period where we have coverage for both exchanges in the pair. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	Log(1+TradeVol)		Log(10+TradeVol)		Log(100+TradeVol)		Log(1000+TradeVol)		Log(10000+TradeVol)	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Listing (0-30 days)	-1.1	0.15	-1.3**	-0.10	-1.4***	-0.32	-1.4***	-0.49**	-1.2***	-0.58***
	(0.73)	(0.35)	(0.61)	(0.30)	(0.49)	(0.25)	(0.39)	(0.20)	(0.30)	(0.15)
Listing $(0-30 \text{ days}) \times \text{Correlation}$	2.1^{**}	0.74^{*}	2.2^{***}	0.98^{***}	2.3^{***}	1.2^{***}	2.2^{***}	1.3***	1.8^{***}	1.3^{***}
	(0.82)	(0.41)	(0.68)	(0.35)	(0.56)	(0.29)	(0.45)	(0.23)	(0.34)	(0.18)
Listing $(> 30 \text{ days})$	-1.3	0.68	-1.7**	0.38	-1.9***	0.11	-1.8***	-0.11	-1.5***	-0.27
	(0.90)	(0.42)	(0.77)	(0.35)	(0.65)	(0.29)	(0.53)	(0.24)	(0.41)	(0.18)
Listing $(> 30 \text{ days}) \times \text{Correlation}$	2.6^{**}	0.29	2.9^{***}	0.55	3.0^{***}	0.76^{**}	2.7^{***}	0.90^{***}	2.2^{***}	0.92^{***}
	(1.0)	(0.47)	(0.87)	(0.40)	(0.73)	(0.34)	(0.60)	(0.27)	(0.47)	(0.21)
Pre Three-day Listing	-1.3*	0.35	-1.3**	0.01	-1.2**	-0.29	-1.1***	-0.54**	-0.91***	-0.69***
	(0.76)	(0.40)	(0.64)	(0.33)	(0.53)	(0.27)	(0.42)	(0.22)	(0.32)	(0.17)
Pre Three-day Listing \times Correlation	2.0**	0.77^{*}	2.0***	1.1***	1.9***	1.4***	1.7^{***}	1.6^{***}	1.4***	1.6***
	(0.83)	(0.45)	(0.71)	(0.38)	(0.59)	(0.32)	(0.47)	(0.26)	(0.36)	(0.22)
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exchange FE-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.85	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.85	0.84
Observations	4,864,077	$8,\!975,\!799$	$4,\!864,\!077$	$8,\!975,\!799$	$4,\!864,\!077$	8,975,799	$4,\!864,\!077$	$8,\!975,\!799$	$4,\!864,\!077$	8,975,799

Table A.7: Central Exchange Listings and Volume: Balanced Panel

This table shows estimates from Specification (17), (18) and (21), where we use a subsample only keeping liquid coins (coin-exchange pairs), which have positive volume each day within a 60-day window around the central exchange's listing date. The dependent variables are log arbitrage volume and trade volume. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin (coin-exchange pair) and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan.

Dependent Variables:	$\log(\text{ArbVol})$		log(TradeVol)			
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Listing (0-30 days)	2.3***	2.0^{***}	0.74^{***}	0.71^{***}	0.12	0.03
	(0.55)	(0.30)	(0.08)	(0.05)	(0.87)	(0.50)
Listing $(> 30 \text{ days})$	2.6^{***}	1.9^{***}	0.57^{***}	0.50^{***}	-2.3	-0.26
	(0.36)	(0.42)	(0.12)	(0.06)	(1.5)	(0.65)
Pre Three-day Listing	1.4^{*}	1.8^{***}	0.68^{***}	1.0^{***}	0.74	0.28
	(0.65)	(0.35)	(0.09)	(0.08)	(0.94)	(0.50)
Listing $(0-30 \text{ days}) \times \text{Correlation}$					0.66	0.78
					(0.96)	(0.56)
Listing $(> 30 \text{ days}) \times \text{Correlation}$					3.0^{*}	0.87
					(1.6)	(0.71)
Pre Three-day Listing \times Correlation					-0.06	0.82
					(1.0)	(0.57)
Coin FE	Yes	Yes	No	No	No	No
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Coin-Exchange Pair FE	No	No	Yes	Yes	Yes	Yes
Exchange FE \times Day FE	No	No	Yes	Yes	Yes	Yes
Standard-Errors	Coin FE & Day FE		Coin-Exchange Pair FE			Day FE
Adjusted \mathbb{R}^2	0.50	0.45	0.76	0.81	0.76	0.81
Observations	$11,\!946$	$25,\!610$	$746,\!141$	$1,\!440,\!781$	$746,\!141$	$1,\!440,\!673$

This table shows estimates from Specification:

$$\begin{aligned} Attention_{c,t} = &\beta_1 PreListing(33-93 \text{ days})_{c,t} + \beta_2 PreListing(3-33 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} \\ &+ \beta_4 Listing(0-30 \text{ days})_{c,t} + \beta_5 Listing(> 30 \text{ days})_{c,t} + \delta_c + \eta_t + \kappa_{matchid,t} + \epsilon_{c,t} \end{aligned}$$

The dependent variable is the log of 1 plus the search volume index. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreListing(33-93 \text{ days})_{c,t}$, $PreListing(3-33 \text{ days})_{c,t}$, and $PreThreedayListing_{c,t}$ are dummy variables which are equal to one for coin *i* on date *t* if a central exchange decides to list coin *i* between date t + 33 and t + 93, t + 3 and t + 33, and t + 1 and date t + 3, respectively. Column (1) is result based on listings on Binance. Column (2) is result based on listings on Coinbase. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Google Search.

Dependent Variables:	Attention			
	Binance	Coinbase		
Model:	(1)	(2)		
Pre Listing (93-33 days)	0.0006	0.06^{*}		
	(0.03)	(0.04)		
Pre Listing (33-3 days)	0.04	0.14^{***}		
	(0.04)	(0.05)		
Pre Listing (1-3 days)	0.09**	0.22***		
	(0.04)	(0.06)		
Listing $(0-30 \text{ days})$	0.16***	0.25***		
	(0.05)	(0.05)		
Listing $(> 30 \text{ days})$	0.06	0.09^{*}		
	(0.04)	(0.05)		
Coin FE	Yes	Yes		
Day FE	Yes	Yes		
Adjusted \mathbb{R}^2	0.86	0.86		
Observations	$1,\!333,\!665$	$1,\!333,\!665$		
This table shows estimates from specification (17):

$$\begin{split} Y_{c,t} = & \beta_1 Listing(0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayListing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. Each treated coin is matched with a control coin based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan, Google Search.

Dependent Variables:	Dispersion		I(Arb)	Vol > 0)	Log(A	rbVol)	Net Listings	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	-0.03***	-0.02**	0.36***	0.22***	3.1***	1.8***	0.09***	0.06***
	(0.009)	(0.008)	(0.09)	(0.04)	(0.46)	(0.33)	(0.01)	(0.01)
Listing $(> 30 \text{ days})$	-0.06***	-0.03**	0.25^{***}	0.18^{***}	1.6^{***}	0.66**	0.004	0.001
	(0.01)	(0.01)	(0.09)	(0.04)	(0.46)	(0.30)	(0.002)	(0.003)
Pre Three-day Listing	0.004	-0.003	0.09	0.005	0.99	1.1***	0.02	0.04^{*}
· · ·	(0.01)	(0.009)	(0.11)	(0.06)	(0.87)	(0.38)	(0.02)	(0.02)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched Coin-Exchange Pair FE-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted \mathbb{R}^2	0.62	0.53	0.36	0.29	0.33	0.43	0.04	0.17
Observations	427,957	382,756	$146,\!362$	$163,\!350$	92,098	119,114	$648,\!195$	457,367

This table shows estimates from Specification (18):

$$Y_{c,e,t} = \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$

The dependent variables are trade volume dummy and log trade volume. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. Each treated coin-exchange pair is matched with a control pair based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Google Search.

Dependent Variables:		I(Trade	$\operatorname{Vol} > 0$			Log(Tr	adeVol)	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	0.01	0.01^{*}	0.02***	0.008**	0.77***	0.56^{***}	0.83***	0.63***
	(0.010)	(0.006)	(0.007)	(0.004)	(0.09)	(0.06)	(0.08)	(0.05)
Listing $(> 30 \text{ days})$	0.04***	0.01^{**}	0.04***	0.007^{*}	0.81^{***}	0.50^{***}	0.87^{***}	0.60***
	(0.01)	(0.007)	(0.009)	(0.004)	(0.13)	(0.08)	(0.11)	(0.06)
Pre Three-day Listing	0.007	0.02***	0.01	0.02***	0.68***	0.83***	0.73***	0.89***
	(0.01)	(0.007)	(0.008)	(0.004)	(0.10)	(0.09)	(0.10)	(0.08)
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exchange FE-Day FE	No	No	Yes	Yes	No	No	Yes	Yes
Adjusted \mathbb{R}^2	0.46	0.47	0.63	0.75	0.74	0.75	0.79	0.83
Observations	1,719,653	$4,\!145,\!078$	1,719,653	$4,\!145,\!078$	$1,\!438,\!901$	$3,\!569,\!786$	$1,\!438,\!901$	$3,\!569,\!786$

Table A.11: Coin-Level Regressions, Attention-Matched Sample (Attention Controls)

This table shows estimates from specification (17):

$$\begin{split} Y_{c,t} = & \beta_1 Listing (0\text{--}30 \text{ days})_{c,t} + \beta_2 Listing (> 30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \beta_4 Attention_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. Each treated coin is matched with a control coin based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Google Search.

Dependent Variables:	Dispersion		I(ArbV	Vol > 0)	Log(A	rbVol)	Net Listings	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	-0.03***	-0.02**	0.35^{***}	0.22^{***}	3.0***	1.5***	0.09***	0.06***
	(0.009)	(0.008)	(0.09)	(0.04)	(0.50)	(0.33)	(0.01)	(0.01)
Listing $(> 30 \text{ days})$	-0.06***	-0.03**	0.25^{***}	0.18^{***}	1.6^{***}	0.59^{*}	0.003	0.0004
	(0.01)	(0.01)	(0.09)	(0.05)	(0.50)	(0.32)	(0.002)	(0.003)
Pre Three-day Listing	0.004	-0.004	0.08	0.002	0.95	1.0^{***}	0.02	0.04^{*}
	(0.01)	(0.009)	(0.11)	(0.06)	(0.92)	(0.39)	(0.02)	(0.02)
Attention	0.0010	0.005	0.05	0.02	0.80***	0.63^{***}	0.004^{***}	0.02***
	(0.005)	(0.004)	(0.03)	(0.02)	(0.14)	(0.19)	(0.002)	(0.004)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched Coin-Exchange Pair FE-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted \mathbb{R}^2	0.62	0.53	0.36	0.29	0.34	0.45	0.04	0.17
Observations	$427,\!957$	382,756	$146,\!362$	$163,\!350$	92,098	$119,\!114$	$648,\!195$	$457,\!367$

Table A.12: Coin-Exchange Level Regressions, Attention-Matched Sample (Attention Controls)

This table shows estimates from Specification (18):

$$\begin{aligned} Y_{c,e,t} = & \beta_1 Listing (0-30 \text{ days})_{c,t} + \beta_2 Listing (> 30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \beta_4 Attention_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{aligned}$$

The dependent variables are trade volume dummy and log trade volume. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. Each treated coin-exchange pair is matched with a control pair based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Google Search.

Dependent Variables:		I(Trade	$\operatorname{Vol} > 0$)		Log(TradeVol)				
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Listing (0-30 days)	0.01	0.01^{*}	0.02***	0.007^{*}	0.72^{***}	0.47^{***}	0.79^{***}	0.54^{***}	
	(0.010)	(0.006)	(0.007)	(0.004)	(0.09)	(0.06)	(0.08)	(0.05)	
Listing $(> 30 \text{ days})$	0.04^{***}	0.01^{*}	0.04^{***}	0.007	0.79^{***}	0.47^{***}	0.85^{***}	0.57^{***}	
	(0.01)	(0.007)	(0.009)	(0.004)	(0.13)	(0.08)	(0.11)	(0.06)	
Pre Three-day Listing	0.007	0.02^{***}	0.01	0.02***	0.67^{***}	0.74^{***}	0.72^{***}	0.80***	
	(0.01)	(0.007)	(0.008)	(0.004)	(0.10)	(0.08)	(0.10)	(0.07)	
Attention	0.003	0.004	-0.002	0.007***	0.46^{***}	0.56^{***}	0.43^{***}	0.55^{***}	
	(0.005)	(0.004)	(0.004)	(0.003)	(0.07)	(0.04)	(0.06)	(0.03)	
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Exchange FE-Day FE	No	No	Yes	Yes	No	No	Yes	Yes	
Adjusted \mathbb{R}^2	0.46	0.47	0.63	0.75	0.74	0.76	0.79	0.83	
Observations	1,719,653	$4,\!145,\!078$	1,719,653	$4,\!145,\!078$	$1,\!438,\!901$	$3,\!569,\!786$	$1,\!438,\!901$	$3,\!569,\!786$	

This table shows estimates from Specification:

$$log (TradeVol_{c,e,t}) = \beta_1 Attention_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \kappa_{matchid,t} + \epsilon_{c,t}$$

The dependent variable is coin-exchange-day level log trade volume. $Attention_{c,t}$ denotes the log of 1 plus the search volume index for coin c at day t. Each treated coin is matched with a control coin based on pre-listing volume levels and attention patterns, as described in Appendix H.3.4. Columns (1) and (2) are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Google Search.

Dependent Variables:	Log(TradeVol)		
	Binance	Coinbase	
Model:	(1)	(2)	
log_SVI	0.44***	0.59^{***}	
	(0.06)	(0.02)	
Coin-Exchange Pair FE	Yes	Yes	
Day FE	Yes	Yes	
Exchange FE \times Day FE	Yes	Yes	
matched_pair_id-Day FE	Yes	Yes	
Adjusted \mathbb{R}^2	0.78	0.84	
Observations	$1,\!438,\!901$	$3,\!569,\!786$	

Table A.14:	Attention	Matched	Sample	Summary	Statistics	

This table shows the summary statistics for the attention-matching balanced data. The first two columns show results based on balanced coin-level data, while the last two columns show results based on balanced coin-exchange-level data. Abs Diff Trend, Diff Level, and Correlation represent the average absolute difference in trade volume between treated and control coins (coin-exchange pairs), the average difference in trade volume between treated and control coins (coin-exchange pairs), and the average correlation in attention between treated and control coins (coin-exchange pairs), respectively. #Matched, #Treated Candidates, and Matching Rate denote the number of successfully matched coins (or coin-exchange pairs), the number of treated coins (or coin-exchange pairs), and the matching rate. Data source: Cryptotick, Google Search.

	Coin	Level	Coin-exch	nange Level
	Binance	Coinbase	Binance	Coinbase
Abs Diff Level	0.95	1.1	1.02	1.01
Diff Level	-0.18	-0.27	-0.16	-0.2
Correlation	0.77	0.79	0.72	0.77
# Matched	163	126	714	1816
# Treated Candidates	182	139	1129	2237
Matching Rate	0.9	0.91	0.63	0.81

Table A.15: Coin Level Regressions and the Endorsement Channel

This table shows estimates from specification (17):

$$\begin{split} Y_{c,t} = & \beta_1 Listing (0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing (> 30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. We restrict our sample to treated coins with a market capitalization of more than \$100 million one month prior to their central exchange listing date. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For price dispersion and arbitrage volume, we exclude newly entered exchanges. For coin listings, we include only coins that were listed on any exchange for at least 30 days before being listed on central exchanges or were never listed by central exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	Dispersion		I(ArbVol > 0)		Log(ArbVol)		Net Listings	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	-0.03**	-0.02**	0.35***	0.19^{***}	1.8***	1.8^{***}	0.04***	0.06***
	(0.01)	(0.009)	(0.07)	(0.04)	(0.29)	(0.29)	(0.007)	(0.01)
Listing $(> 30 \text{ days})$	-0.06***	-0.03***	0.37^{***}	0.21^{***}	2.0^{***}	1.6^{***}	0.01^{**}	0.007
	(0.01)	(0.01)	(0.06)	(0.05)	(0.32)	(0.33)	(0.004)	(0.006)
Pre Three-day Listing	-0.02	-0.01	0.06	0.14^{***}	0.74^{**}	1.6^{***}	0.002	0.06^{**}
	(0.01)	(0.008)	(0.09)	(0.04)	(0.35)	(0.30)	(0.009)	(0.03)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.51	0.29	0.41	0.39	0.49	0.50	0.17	0.21
Observations	$124,\!939$	$140,\!805$	$39,\!971$	$65,\!382$	$27,\!681$	50,524	$168,\!605$	$157,\!484$

This table shows estimates from Specification (18):

$$Y_{c,e,t} = \beta_1 Listing(0-30 \text{ days})_{c,t} + \beta_2 Listing(> 30 \text{ days})_{c,t} + \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t}$$

The dependent variables are trade volume dummy and log trade volume. We restrict our sample to treated coins with a market capitalization of more than 100 million one month prior to their central exchange listing date. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listing 1-2 days in advance. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:		I(Trade	$\operatorname{Vol} > 0$		Log(TradeVol)				
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Listing (0-30 days)	0.02^{*}	0.007	0.02***	0.005	0.63***	0.54^{***}	0.61***	0.59***	
	(0.01)	(0.006)	(0.007)	(0.004)	(0.09)	(0.06)	(0.08)	(0.05)	
Listing $(> 30 \text{ days})$	0.04^{***}	0.01	0.04^{***}	0.004	0.71^{***}	0.42^{***}	0.53^{***}	0.51^{***}	
	(0.02)	(0.008)	(0.01)	(0.005)	(0.13)	(0.06)	(0.13)	(0.06)	
Pre Three-day Listing	0.008	0.02***	0.01^{*}	0.01***	0.47^{***}	0.87***	0.52^{***}	0.91***	
	(0.01)	(0.006)	(0.008)	(0.004)	(0.12)	(0.09)	(0.10)	(0.08)	
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Exchange $FE \times Day FE$	No	No	Yes	Yes	No	No	Yes	Yes	
Adjusted \mathbb{R}^2	0.49	0.49	0.70	0.77	0.74	0.77	0.80	0.84	
Observations	878,830	$2,\!175,\!539$	878,830	$2,\!175,\!539$	$727,\!869$	$1,\!843,\!344$	$727,\!869$	$1,\!843,\!344$	

This table shows estimates from Specification (18):

$$\begin{split} Y_{c,e,t} = & \beta_1 Listing (0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing (>30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variables are trade volume dummy and log trade volume. We keep 59 exchanges that have both AML and KYC requirements. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:		I(Trade	$\operatorname{Vol} > 0$		Log(TradeVol)				
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Listing (0-30 days)	0.02	0.02***	0.03***	0.02^{**}	0.89^{***}	0.98***	1.0^{***}	0.99***	
	(0.01)	(0.008)	(0.009)	(0.006)	(0.13)	(0.09)	(0.12)	(0.08)	
Listing $(> 30 \text{ days})$	0.04^{**}	0.02^{**}	0.05^{***}	0.02^{***}	0.83^{***}	1.0^{***}	1.0^{***}	1.1^{***}	
	(0.02)	(0.009)	(0.01)	(0.007)	(0.18)	(0.11)	(0.16)	(0.09)	
Pre Three-day Listing	0.02	0.03***	0.03^{**}	0.02***	0.79^{***}	1.2^{***}	0.90***	1.3^{***}	
	(0.01)	(0.009)	(0.01)	(0.006)	(0.14)	(0.11)	(0.13)	(0.10)	
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Exchange FE \times Day FE	No	No	Yes	Yes	No	No	Yes	Yes	
Adjusted \mathbb{R}^2	0.54	0.56	0.71	0.75	0.72	0.72	0.77	0.76	
Observations	$1,\!394,\!050$	$2,\!805,\!365$	$1,\!394,\!050$	$2,\!805,\!365$	$1,\!153,\!621$	$2,\!359,\!891$	$1,\!153,\!621$	$2,\!359,\!891$	

Table A.18: Coin-Exchange Level Regressions and Fake Volume: Intensive Margin

This table shows estimates from Specification (18):

$$\begin{split} \log \left(TradeVol_{c,e,t} \right) = & \beta_1 Listing(0\text{--}30 \text{ days})_{c,t} + \beta_2 Listing(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variable is log trade volume. We keep 29 exchanges analyzed in Cong et al. (2023b). $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For Columns (1) and (2), we keep 3 exchanges in the regulated group. For Columns (3) and (4), we keep 10 exchanges in the Tier 1 group. For Columns (5) and (6), we keep 16 exchanges in the Tier 2 group. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	Log(TradeVol)								
	Regi	ulated	Tie	er 1	Ti	er 2			
	Binance	Coinbase	Binance Coinbase		Binance	Coinbase			
Model:	(1)	(2)	(3)	(4)	(5)	(6)			
Listing (0-30 days)	0.56	0.46^{*}	0.93***	0.96***	1.1^{***}	0.83***			
	(0.41)	(0.23)	(0.11)	(0.08)	(0.20)	(0.10)			
Listing $(> 30 \text{ days})$	0.07	0.09	0.92^{***}	1.1^{***}	1.1^{***}	0.82^{***}			
	(0.37)	(0.26)	(0.15)	(0.10)	(0.21)	(0.11)			
Pre Three-day Listing	0.28	0.95^{***}	0.89^{***}	1.3^{***}	0.80^{***}	0.93^{***}			
	(0.35)	(0.19)	(0.13)	(0.11)	(0.18)	(0.11)			
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes			
Day FE	Yes	Yes	Yes	Yes	Yes	Yes			
Exchange FE \times Day FE	Yes	Yes	Yes	Yes	Yes	Yes			
Adjusted \mathbb{R}^2	0.90	0.89	0.77	0.77	0.77	0.76			
Observations	$55,\!081$	35,826	$1,\!202,\!135$	$2,\!310,\!313$	$475,\!683$	$929,\!571$			

Table A.19: Coin-Exchange Level Regressions and Fake Volume: Extensive Margin

This table shows estimates from Specification (18):

$$\begin{split} I\left(TradeVol_{c,e,t} > 0\right) = & \beta_1 Listing(0\text{--}30 \text{ days})_{c,t} + \beta_2 Listing(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThreedayListing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variable is trade volume dummy. We keep 29 exchanges analyzed in Cong et al. (2023b). $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For Columns (1) and (2), we keep 3 exchanges in the regulated group. For Columns (3) and (4), we keep 10 exchanges in the Tier 1 group. For Columns (5) and (6), we keep 16 exchanges in the Tier 2 group. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	I(TradeVol > 0)								
	Regi	Regulated Tier 1			Ti	er 2			
	Binance	Coinbase	Binance Coinbase		Binance	Coinbase			
Model:	(1)	(2)	(3)	(4)	(5)	(6)			
Listing (0-30 days)	0.11	-0.04	0.02**	0.02***	0.01	0.008			
	(0.08)	(0.03)	(0.010)	(0.007)	(0.01)	(0.008)			
Listing $(> 30 \text{ days})$	0.11	0.04	0.05***	0.04^{***}	0.02^{*}	0.01			
	(0.09)	(0.03)	(0.01)	(0.008)	(0.01)	(0.009)			
Pre Three-day Listing	0.11	-0.05	0.02**	0.02^{***}	0.004	0.02^{*}			
	(0.08)	(0.05)	(0.009)	(0.007)	(0.02)	(0.009)			
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes			
Day FE	Yes	Yes	Yes	Yes	Yes	Yes			
Exchange FE \times Day FE	Yes	Yes	Yes	Yes	Yes	Yes			
Adjusted \mathbb{R}^2	0.71	0.41	0.58	0.62	0.79	0.82			
Observations	$57,\!185$	38,239	$1,\!372,\!158$	$2,\!578,\!831$	$638,\!491$	$1,\!239,\!375$			

This table shows estimates from Specification (17):

$$\begin{split} Y_{c,t} = & \beta_1 Listing (0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing (> 30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. The sample period ranges from January 2017 to July 2022. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For price dispersion and arbitrage volume, we exclude newly entered exchanges. For coin listings, we include only coins that were listed on any exchange for at least 30 days before being listed on central exchanges or were never listed by central exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan.

Dependent Variables:	Dispersion		I(ArbVol > 0)		Log(ArbVol)		Net Listings	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	-0.04***	-0.04***	0.38^{***}	0.29***	2.0***	2.1^{***}	0.06^{***}	0.08***
	(0.01)	(0.009)	(0.05)	(0.04)	(0.28)	(0.25)	(0.01)	(0.01)
Listing $(> 30 \text{ days})$	-0.08***	-0.07***	0.39***	0.32^{***}	1.8^{***}	1.9^{***}	0.01^{***}	-0.0006
	(0.01)	(0.01)	(0.05)	(0.04)	(0.23)	(0.32)	(0.003)	(0.005)
Pre Three-day Listing	-0.02	-0.03***	0.13^{**}	0.23^{***}	1.2^{***}	1.4^{***}	0.02	0.05^{*}
	(0.01)	(0.009)	(0.06)	(0.04)	(0.30)	(0.26)	(0.01)	(0.03)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.52	0.49	0.43	0.42	0.58	0.55	0.06	0.06
Observations	$517,\!305$	$659,\!892$	$144,\!566$	181,764	83,033	$110,\!049$	$946,\!146$	$1,\!026,\!227$

This table shows estimates from Specification (17):

$$\begin{split} Y_{c,t} = & \beta_1 Listing(0\text{--}30 \text{ days})_{c,t} + \beta_2 Listing(>30 \text{ days})_{c,t} + \\ & \beta_3 PreThree dayListing_{c,t} + \delta_c + \eta_t + \epsilon_{c,t} \end{split}$$

The dependent variables are price dispersion, arbitrage volume dummy, log arbitrage volume, and the change in the number of exchanges. Price dispersion is calculated as the standard deviation of log prices across exchanges. The sample period ranges from August 2022 to December 2023. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. For price dispersion and arbitrage volume, we exclude newly entered exchanges. For coin listings, we include only coins that were listed on any exchange for at least 30 days before being listed on central exchanges or were never listed by central exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick, Etherscan.

Dependent Variables:	Dispersion		I(ArbVol > 0)		Log(ArbVol)		Net Listings	
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	-0.03	-0.005	0.74^{***}	0.17^{***}	3.4^{***}	2.0***	0.21^{***}	0.07**
	(0.02)	(0.01)	(0.08)	(0.06)	(0.33)	(0.50)	(0.05)	(0.03)
Listing $(> 30 \text{ days})$	-0.06	-0.007	0.70^{***}	0.21^{***}	2.6^{***}	1.6^{***}	0.004	0.002
	(0.04)	(0.01)	(0.08)	(0.06)	(0.25)	(0.49)	(0.008)	(0.007)
Pre Three-day Listing	-0.009	0.03	0.14	0.06	-0.19	2.4^{***}	0.06	0.02
	(0.02)	(0.02)	(0.10)	(0.08)	(0.81)	(0.67)	(0.04)	(0.03)
Coin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted \mathbb{R}^2	0.74	0.74	0.49	0.47	0.54	0.55	0.11	0.12
Observations	$128,\!041$	$187,\!529$	$55,\!478$	$74,\!567$	33,701	46,529	$278,\!332$	$314,\!925$

Table A.22: Central Exchange Listings: Coin-Exchange Level (Jan 2017 to July 2022)

This table shows estimates from Specification (18):

$$\begin{split} Y_{c,e,t} = & \beta_1 Listing (0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing (>30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variables are trade volume dummy and log trade volume. The sample period ranges from January 2017 to July 2022. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	I(TradeVol > 0)				Log(TradeVol)			
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	0.01	0.02***	0.02***	0.02***	0.64***	0.91***	0.70***	0.89***
	(0.008)	(0.006)	(0.006)	(0.004)	(0.07)	(0.06)	(0.06)	(0.05)
Listing $(> 30 \text{ days})$	0.04^{***}	0.05^{***}	0.04^{***}	0.04^{***}	0.71^{***}	1.1^{***}	0.81^{***}	1.1^{***}
	(0.010)	(0.007)	(0.007)	(0.004)	(0.10)	(0.07)	(0.08)	(0.05)
Pre Three-day Listing	0.004	0.03***	0.006	0.02***	0.56^{***}	1.1^{***}	0.62^{***}	1.1***
	(0.009)	(0.006)	(0.007)	(0.004)	(0.09)	(0.08)	(0.08)	(0.08)
Coin-Exchange Pair FE	Yes							
Day FE	Yes							
Exchange FE \times Day FE	No	No	Yes	Yes	No	No	Yes	Yes
Adjusted \mathbb{R}^2	0.46	0.48	0.62	0.68	0.80	0.79	0.85	0.85
Observations	3,766,644	$6,\!536,\!601$	$3,\!766,\!644$	$6,\!536,\!601$	$2,\!983,\!936$	$5,\!321,\!089$	$2,\!983,\!936$	$5,\!321,\!089$

Table A.23: Central Exchange Listings: Coin-Exchange Level (Aug 2022 to Dec 2023)

This table shows estimates from Specification (18):

$$\begin{split} Y_{c,e,t} = & \beta_1 Listing (0\text{-}30 \text{ days})_{c,t} + \beta_2 Listing (>30 \text{ days})_{c,t} + \\ & \beta_3 PreThree day Listing_{c,t} + \delta_{c,e} + \eta_t + \gamma_{e,t} + \epsilon_{c,e,t} \end{split}$$

The dependent variables are trade volume dummy and log trade volume. The sample period ranges from August 2022 to December 2023. $Listing(0-30 \text{ days})_{c,t}$ and $Listing(> 30 \text{ days})_{c,t}$ are dummies for coins listed within the past 30 days or more than 30 days ago. $PreThreedayListing_{c,t}$ accounts for the pre-announcement effect, given that Binance and Coinbase typically announce listings 1-2 days in advance. We keep only those incumbent coin-exchange pairs that were listed at least 30 days before their listing on central exchanges, or pairs with coins that have not been listed by central exchanges, in order to identify the listing effect on incumbent exchanges. Odd and even columns are results based on listings on Binance and Coinbase, respectively. Standard errors are clustered at the coin-exchange pair and time level. Standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels respectively. Data source: Cryptotick.

Dependent Variables:	I(TradeVol > 0)				Log(TradeVol)			
	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase	Binance	Coinbase
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing (0-30 days)	0.09***	0.006	0.06***	0.0004	1.5^{***}	0.24^{***}	1.5^{***}	0.25***
	(0.02)	(0.009)	(0.02)	(0.007)	(0.20)	(0.07)	(0.17)	(0.07)
Listing $(> 30 \text{ days})$	0.10***	0.004	0.06**	0.002	1.2^{***}	0.20**	1.1^{***}	0.20***
	(0.03)	(0.009)	(0.03)	(0.006)	(0.26)	(0.08)	(0.26)	(0.07)
Pre Three-day Listing	0.07***	0.008	0.04**	0.008	1.3^{***}	0.62^{***}	1.4^{***}	0.61***
	(0.02)	(0.01)	(0.02)	(0.008)	(0.26)	(0.21)	(0.26)	(0.19)
Coin-Exchange Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exchange $FE \times Day FE$	No	No	Yes	Yes	No	No	Yes	Yes
Adjusted \mathbb{R}^2	0.64	0.65	0.75	0.77	0.87	0.85	0.89	0.87
Observations	808,836	$1,\!832,\!801$	808,836	$1,\!832,\!801$	$561,\!048$	$1,\!378,\!958$	$561,\!048$	$1,\!378,\!958$