Limit order book spreads, depth, and market efficiency in a general equilibrium model

Scott Condie* Brennan Platt†‡

ABSTRACT

This paper studies equilibrium order book formation in a limit-order market, building a search-theoretic model driven by participants' expectations about the arrival of liquidity takers to the market. Equilibrium pricing is characterized in terms of competitive balance, which endogenously splits the gains from trade between buyers and sellers. This determines the traditional measures of market liquidity (spread and price impact) and plays a major role in expected welfare. We characterize the determinants of welfare and find that greater liquidity does not always increase market welfare. Additionally, the paper demonstrates how to extract the model's parameters from order book data and applies these techniques to a sample of orderbook data from Coinbase's Bitcoin/U.S. Dollar exchange. We estimate that the exchange generates 83% of the welfare that would be possible in a frictionless market on average over time, but that the estimated welfare efficiency of the market varies over time and is highly skewed. We calculate the variation in spread and in welfare over time and decompose it into differences in participant valuations, differences in beliefs about liquidity taker arrival rates and differences in the likelihood of liquidity takers finding a price at which they are willing to transact.

Keywords: Order book, spread, price impact, equilibrium search, liquidity. market microstructure, welfare, Bitcoin

JEL classification: G14, D53

*Department of Economics, Brigham Young University. Email: ssc@byu.edu. Telephone: +1.801.422.5306.

†Department of Economics, Brigham Young University. Email: bcplatt@byu.edu. Telephone +1.801.422.8904.

‡The authors appreciate insightful feedback from Bradley Larsen, Joey McMurray, John Stovall, Pierre-Olivier Weill and Jonathan Brogaard.

1 Introduction

In most equity markets throughout the world, market makers provide liquidity to those wishing to purchase or sell assets by placing limit orders. These limit orders remain on the book until either executed or canceled. An execution occurs when a participant places a market order, which transacts at the most favorable price on the order book.

This arrangement leaves each side of the market with distinct uncertainty. Those who place limit orders choose their exact price, but must await the arrival of enough market orders to reach that price. Meanwhile, those who place market orders are expressing a desire to transact imediately; yet they face uncertainty in the transaction price if several market orders occur in close succession. Moreover, changing market conditions could undermine either side's motivation to trade.

This paper presents a model of order book trading where these uncertainties are all rooted in the random arrival rate of liquidity providers (those placing limit orders) and liquidity demanders (those placing market orders). For ease of exposition, we focus on the ask side of the market and refer to liquidity providers as sellers and liquidity demanders as buyers. The bid side of the market would proceed similarly. In this model, sellers understand that a higher price will only transact when buyers are especially plentiful, and buyers anticipate that their transaction price depends on how many other buyers arrive near the same time. We depict their decision making in a dynamic search environment.

Traders enter the model knowing their valuation of the asset.[1] Sellers share a common view of the asset's fundamental value, which is strictly less than the valuation of all buyers. Thus, trade is always beneficial, which the order book facilitates by determining the terms of trade. Both sides of the market may postpone participation and not place an order if it is optimal to do so. Despite the homogeneity of participants, the order book always generates dispersion in the asset price, without requiring the presence of noise traders who would transact regardless of the prices they face.

This paper demonstrates that the net effect of participating traders' behavior on the equilibrium order book can be characterized in terms of *competitive balance*, which

[1]That is, we explore the extent to which order book prices can be explained from variation in participant arrivals rather than variation in participant valuations.

measures the fraction of surplus that a seller receives from the average transaction. This metric is derived from the effective discount rate and the joint distribution of buyer and seller arrivals. While one might expect that the competitive balance always shifts towards buyers as sellers are more plentiful, this is not always the case. In particular, buyers are better off when the arrival rates are more similar on both sides, while sellers benefit from greater difference in the arrival rates.

This relates to the broader concept of *market liquidity* that is used in practice to characterize how easily one can transact in an order book market. Greater liquidity manifests in a narrower spread (*i.e.* the difference between the highest bid limit order and the lowest ask limit order) or in a smaller price impact (*i.e.* if buying X shares, the difference between the first share's price and the X^{th} share's price). In the literature, there is evidence that a lack of liquidity is perceived by traders to be risky in the aggregate (Pástor and Stambaugh, 2003), causes higher fees associated with seasoned equity offerings (Butler et al., 2005), and leads to less informative prices (Kerr et al., 2020). In our model, the market becomes more liquid as the competitive balance shifts towards buyers. Thus, more liquid markets are better for buyers, but worse for sellers.

Since we fully model trader preferences, we can meaningfully evaluate the expected welfare gains associated with the market, and find that greater liquidity does not necessarily imply greater welfare. In some cases, welfare and liquidity move in tandem (such as when traders are more patient or market participants have greater parity in arrival rates), but these concepts can also diverge. For example, when the joint surplus of each transaction is larger, it will proportionally scale up expected welfare, spread, and price impact. Moreover, a shift in competitive balance away from sellers will always improve liquidity, but can eventually reduce welfare if it leads to many unmatched sellers. These welfare consequences are presented theoretically as well as quantitatively for the sample of data that we analyze.

To illustrate the model's potential explanatory power, we apply it to order book data from Coinbase's Bitcoin/U.S. Dollar exchange.[2] Our calibration process interprets the data through the lens of our model. First, we parse the flow of updates to the order book into time periods (regimes) in which the best ask and best bid prices do not cross during the regime. Then we provide an estimation technique that

[2]These data are convenient to collect and of independent interest, but the model is equally applicable to any limit-order market.

estimates the average order book in each regime, while accounting for bias that naturally arises when the lowest price is executed first and thus is frequently unobserved. In this extraction process, we leverage the structure of the model, using the equal profitability of various limit orders to determine implied beliefs about the arrival of market participants. Compared across regimes, this sheds light on the underlying forces that drive liquidity and welfare.

We find that the average estimated asset valuation difference between buyers and sellers is \$1.51 in these data, with an observed average spread of \$0.16. The competitive balance lies heavily in favor of buyers, as we find that just over 7% of the gains from trade go to the sellers on average. We document the role that implied beliefs about the distribution of trader arrival and differences in trader valuations have on the competitive balance between buyers and sellers, and study the extent to which these predict the observed spread. We find that in this sample approximately 28% of the variation in the spread can be explained by differences in competitive balance, while the remainder is due to variation in the difference between buyer and seller valuations. The market is relatively efficient, delivering on average 83% of the welfare gains that could be obtained from a frictionless market, but this distribution is highly skewed with a median welfare over time of 27%. We also examine what drives the variation in welfare across regimes and decompose this into it's component parts, the difference in trader valuations being the most important determinant.

We see the contribution of this paper as three-fold. The first contribution is that we provide an equilibrium model that makes predictions about both the spread and the price impact in a limit order market. Our analysis is built on the uncertain arrival of participants in the market, which is the fundamental motivation for organizing a market with an order book. Second, our paper allows us to analyze the welfare implications of order books because all traders are modeled explicitly from their preferences. Finally, our paper makes a methodological contribution by demonstrating how to extract beliefs about arrival of participants from typical order book data, and then use these beliefs to evaluate expected welfare.

We proceed in the paper by first reviewing the literature, after which we build a general model in section 2, analyzing competitive balance, liquidity and welfare in equilibrium. In section 3, we illustrate how to apply the general model to observable order book data, assess the fit of the model, and evaluate the welfare gains in this market. Section 4 concludes.

1.1 Literature

We contribute to a robust literature on the formation of order books, starting with the canonical models of Glosten and Milgrom (1985) and Kyle (1985). These both feature some fraction of uninformed noise traders, who must purchase regardless of the market price; market makers do not know whether market orders come from an informed or uninformed trader, but update their beliefs about the asset's value in the market by observing the order flow. The equilibrium competition among market makers determines the spread in Glosten and Milgrom (1985) or the price impact in Kyle (1985).[3] They find that increased competition leads to tighter spreads or flattens the price impact in the orderbook. These are both true in our setting, though we do so without the presence of noise traders, which enables us to examine welfare consequences for the market as a whole.

Easley and O'Hara (1987) models the order book as a game of adverse selection, where buyers have private information about the value of the asset. Uninformed sellers may be able to deduce this hidden information through the size of market orders placed, so the various limit-order prices reflect the changing willingness to sell as order size increases. In Hollifield et al. (2004), traders submit different limit orders because some value the asset more and are willing to accept a lower (exogenously specified) probability of selling it. Our model is similar in that sellers recognize that a higher price is less likely to trade, but probabilities are determined endogenously and adjust so as to make sellers indifferent. This creates a range of equally-profitable limit orders even without differential information.

The model of Hollifield et al. (2004) also includes a minimum tick size for prices, which creates a valuation threshold for whether it is worth purchasing at a price that is one tick higher. They exploit this for a compelling empirical test of whether the model-implied threshold shifts monotonically with the probability of a limit order being executed, which is supported in Ericsson stock data from the Stockholm Automated Exchange (from December 1992 to March 1993).

Hollifield et al. (2006) extends Hollifield et al. (2004) to continuous time and incorporates a cost of executing an order, which they use to investigate welfare from the trade of three mining stocks traded on the Vancouver Stock Exchange (between May 1990 and November 1993). They find that the market realizes 90% of the potential

[3]Back and Baruch (2004) nests both models in a continuous time model and establishes conditions under which equilibrium converges to that of Glosten and Milgrom (1985) and Kyle (1985).

gains from trade from a perfectly liquid market, with inefficiency arising due to non-execution of some limit orders. We find similar quantitative welfare results, although our results are driven by the mismatch in the arrival of participants that prevent some transactions from occurring while they are still socially beneficial. Transaction probabilities are endogenous in our model and the participant arrival distribution is non-parametrically estimated, which appear to be key contributors to our lower welfare estimate.

Another strand of literature has modeled centralized exchange in the presence of search frictions.[4] These build dynamic models of limit-order markets where all traders' preferences are fully modeled, as in our paper. In Goettler et al. (2005), traders can only cancel old limit orders with some exogenous probability. All traders agree on the value of the asset at a given time, but that value can change each period, creating a risk that a limit order is executed when it is no longer beneficial to the trader who offered it. Trade in Foucault et al. (2005) is motivated by differing waiting costs, where low cost traders post limit orders that are executed by market orders from high cost traders. By assumption, new limit orders must be price improving (i.e. narrow the spread) by the exogenous tick amount. Roşu (2009) has a similar setting except that neither a minimum tick nor price improving requirement is imposed; rather, both arise endogenously as the most profitable limit order to add to the order book. Dugast (2018) focuses on changes in investor valuations over time, generating predictions about the relationship between the frequency of news arrival, market depth, and executions.

In the models of all four papers, the order book is dynamically built up over time as arriving traders either place limit orders or execute them in market orders. This generates predictions about low-level trends from specific transactions on the order book: the correlation between transaction costs and spread, the distribution of market spread and execution times, and bid and ask prices display co-movement, respectively. In contrast, our model characterizes the order book in steady state, averaged over short periods where asset fundamentals are steady but arrival of traders may vary. This allows us to examine a more aggregate perspective: the distribution of limit orders informs us of the anticipated arrival rates of participants and thus

[4]Search theory is frequently applied in studying over-the-counter exchange, where buyers and sellers must find each other through a decentralized process. Weill (2020) provides a recent survey, including the relatively sparse application to centralized trade.

reveals expected welfare. Our model also allows free entry and exit from limit orders, and approximates a large market with a continuum of traders. This generates an analytic solution that is particularly useful in deriving empirical implications for the order book that we can take to data.

Several search theoretic papers study liquidity in asset markets from a macro perspective. Lagos and Rocheteau (2009) studies liquidity in over-the-counter markets. This model makes predictions about liquidity measures like the bid-ask spread as in our paper, although they do not attempt to model the mechanics of price priority in limit-order books as is done here. Cui and Radde (2016) also abstracts from the specific properties of the limit-order book, embedding a financial sector with search frictions into a dynamic general equilibrium consumption-saving-investment model. Vayanos and Wang (2007) builds a search-based model of trading with the friction that traders can search for only one asset and then studies the equilibria that result. They show that in one equilibrium, short-horizon investors congregate in one market and that this market is more liquid than the other market with long-horizon investors. The question of time horizon does not enter into our model since all traders have the same trading horizon.

2 An order book model

Consider a market for an asset, with risk neutral sellers and buyers. Sellers are homogenous, each holding one unit of the asset.[5] Upon entering the market, the seller places a limit order, which means he selects a price at which he is willing to sell. The set of limit orders form the order book. Buyers are also homogenous, with demand for one unit of the asset. Upon entering the market, each places a market order, which means she will transact at the lowest available price on the order book. The strategic choice of a buyer is to postpone purchasing if available prices are too high.[6]

Arrivals and transactions are modeled in continuous time. Indeed, the delay (al-

[5]Incorporating multi-unit supply or demand would only be consequential in our model if a single agent's order would significantly alter the order book. In that case, the agent (buyer or seller) may want to strategically spread out the order over time.

[6]This behavior is consistent with buyers' submission of marketable limit orders—that is, limit buy orders that have a price that crosses the spread. Throughout the paper we use the term "market orders" as a short hand for these marketable limit orders.

beit small) between the submission and the posting of orders means that others' orders may be posted in the interim. Thus, agent decisions are not made contingent on the lowest price at that instant (which could rise as more market orders arrive or fall as more limit orders arrive). Rather, agents decide based on their expectation that market transactions will reach at least a price of a, represented by probability $G(a)$. We define $a(0)$ as the lowest expected price, meaning that $G(a) = 1$ for all $a \leq a(0)$, and $a(1)$ as the highest expected price, meaning $G(a)$ is constant for all $a \geq a(1)$ (though not necessarily at $G(a) = 0$). The function $G(a)$ can be interpreted as agents' steady state expectations about the order book. Sellers take these probabilities as given while selecting a limit order price, as do buyers in setting a maximum price for their market order; even so, $G(a)$ is endogenously determined by the sellers' decisions of limit order prices. These choices are made simultaneously.

The realized ratio of buyers to sellers at each instant, denoted q, is a random variable drawn from cumulative distribution $F(q)$ for $q \geq 0$, which we assume to be continuously differentiable for all q, with $F'(q) > 0$ for all $q \leq 1$. When a buyer's market order is posted, the realized q at that moment determines the fraction of shares on the order book that are transacted, and thus what price the buyer pays. Likewise, when a seller's limit order is posted, the sequence of realized qs determines how long the seller must wait before enough shares are transacted to reach the posted price a. This creates uncertainty for buyers about price and for sellers about the time required for transaction.

We also assume that this market takes place on an exchange which coordinates buyer and sellers through the order book. When transactions are realized, the exchange pays the seller a rebate f_m and collects a take fee f_t from the buyer. We take these exchange fees to be exogenous.[7]

2.1 Sellers

Upon entry, each seller places a limit order to sell one unit of the asset. Let $V_S(a)$ denote the seller's present expected value of placing a limit order with ask price a. A seller's flow utility comes from three components.

First, while holding the asset, sellers receive flow value d_A, whether from its dividends or from its perceived value as part of the seller's portfolio.

[7] In practice, the rebate is meant to encourage sellers to provide liquidity, while the take fee can pay for the rebate plus any cost of operating the exchange.

Second, the order may transact if a becomes the lowest price on the order book (*e.g.* all lower prices have transacted). This occurs with probability $G(a)$. Upon execution, the seller is paid a by the buyer as well as the rebate f_m by the exchange, but relinquishes the asset, leading to a net gain of $a + f_m - V_S(a)$.

Finally, at Poisson rate θ, the trading opportunities cease and the seller continues to receive flow value d_A indefinitely.[8] In our continuous time environment with a discount rate ρ, the seller's Bellman function is:

$$\rho V_S(a) = d_A + G(a)(a + f_m - V_S(a)) + \theta \left(\frac{d_A}{\rho} - V_S(a) \right). \tag{1}$$

Sellers can then use this equation to compute the expected value of posting a price a, knowing that a higher price has less chance of transacting and thus implies a longer delay of $1/G(a)$ units of time, on average.

2.2 Population Dynamics

At each instant, a ratio of buyers to sellers q is drawn from the exogenous distribution $F(q)$. These draws are assumed to be independent over time. $F(q)$ represents the exogenous beliefs of agents on both sides about the relative number of shares that will be desired by market orders relative to share on offer on the book. This model does not inspect the rationality of those beliefs; rather, we will use order book data to back out what market makers must believe about those arrival rates.

As an example of such a distribution,[9] suppose that the number of shares n_a that sellers offer as limit orders are exponentially distributed with mean $1/\alpha$, while the number of shares n_b that buyers seek as market orders are independently exponentially distributed with mean $1/\beta$. The joint distribution then becomes $f(n_a, n_b) = \alpha e^{-\alpha n_a} \beta e^{-\beta n_b}$. If we define the ratio of market orders to limit orders as $q = \frac{n_b}{n_a}$ and define $\phi = \frac{\beta}{\alpha}$ (the average number of sellers per buyer), this cumulative density can be transformed to become $F(q) = \frac{\phi q}{1+\phi q}$.

[8] This shock occurs to all buyers and sellers simultaneously. This shock can be interpreted as buyers' information becoming obsolete, removing their interest in buying the asset and leaving the sellers with the asset forever. One could expand the model, depicting this as a shock that initiates a new steady state, with transitions between steady state regimes depicted as a Markov process, but this would have minimal impact on the model predictions with significant notational complication.

[9] Our results do not depend on the functional form in this example, but we will refer to it where useful for intuition.

If the realized ratio generates $q \leq 1$, then all market orders are executed but some limit orders will not be. Indeed, the realized ratio of buyers to sellers q will dictate how much of the orderbook is used up, and therefore implies what the lowest remaining price $a(q)$ must be: namely, the endogenous probability of reaching price $a(q)$ must equal the exogenous probability of having drawn ratio q.

$$G(a(q)) = 1 - F(q) \text{ if } q \leq 1. \tag{2}$$

If the realized ratio $q > 1$, then all of the limit orders on the book are executed, while some market orders will not be executed. Thus, $q < 1$ results in delayed transactions for some sellers, while $q > 1$ results in delayed transactions for some buyers.

2.3 Buyers

We next turn to the buyers' side of the market. Let V_B denote the buyer's present expected value of placing a marketable order with maximum price $\bar{a}$. Note that the buyer could fully abstain from the market by choosing $\bar{a} < a(0)$. After deciding on the market order, it is received by the exchange at Poisson rate of 1, leading to the average delay of 1 units of time (which is the same rate between trading opportunities for the sellers). At rate θ, the trader's information about the asset become obsolete and the buyer's search ends, for a net change of $0 - V_B$.

When the buyer's order is received, the ratio q is realized. If $q \leq 1$, all buyers are able to transact. If $q > 1$, the sales will be rationed among buyers, serving only $1/q$ of them. Upon transacting, the buyer receives a perpetual flow of d_B. We assume that $d_B > d_A$ to ensure that the buyers who arrive want to purchase the asset from sellers.[10] The buyer pays the take fee f_t and the realized price $a(s)$, where s is drawn from a uniform distribution over $[0, \min\{q, 1\}]$. That is, the buyer is equally likely to be in any position in the order arrival process, and that order determines which order-book price the buyer receives. In sum, this yields net utility $\frac{d_B}{\rho} - a(s) - f_t - V_B$.

[10]The difference in flow value between buyers and sellers is fundamental to generating trade. In the analogous bid market, were it to be modeled, those placing limit orders to buy must value the asset more than those placing market orders to sell, but both would have to value the asset less than participants in the ask market.

We can thus compute the value function for a buyer as:

$$\rho V_B = \int_0^1 \left[\int_0^q \frac{1}{q}\left(\max\left\{\frac{d_B}{\rho} - a(s) - f_t - V_B, 0\right\}\right) ds\right] F'(q)dq + \int_1^\infty \left(\int_0^1 \left(\max\left\{\frac{d_B}{\rho} - a(s) - f_t - V_B, 0\right\}\right) ds\right) \frac{1}{q}F'(q)dq + \theta(0 - V_B). \tag{3}$$

The first term in parenthesis handles realizations of $q \leq 1$. Note that the buyer does transact, and that position s is distributed with density $1/q$, determining the price $a(s)$ of that transaction. The second term handles realizations when $q > 1$, so only fraction $1/q$ of buyers transact, while position s is distributed with density 1.

In this setting, buyers passively accept the best price available on the order book, so their only strategic choice is whether to reject the current opportunity (i.e. due to the position s they have drawn) in favor of future opportunities. This results in a reservation price $\bar{a} = \frac{d_B}{\rho} - f_t - V_B$. Any price at or below this threshold will generate a net increase in utility relative to waiting. As an immediate consequence, the value of search must be weakly positive since buyers can set $\bar{a} < a(0)$.

2.4 Equilibrium

New entrants make strategic decisions simultaneously, creating a normal form game. The Nash equilibrium consists of a pricing function $a(q)$ for $q \leq 1$, transaction probabilities $G(a)$, and a reservation price $\bar{a}$ such that:

1. The support of $G(a)$ equals $[a(0), a(1)]$.
2. Every price in the support of $G(a)$ is equally profitable, maximizing Eq. 1, while prices outside the support are weakly less profitable.
3. Prices obey the population dynamics in Eq. 2
4. Buyers are willing to accept all prices in the support, meaning $\bar{a} = a(1)$ is the optimal reservation price for Eq. 3.

Conditions 1 and 4 ensure that all prices offered on the order book can be executed under some circumstance. If Condition 4 failed, sellers would be listing some prices that no one would ever accept, violating Condition 2. Condition 2 adds that sellers must be indifferent across all the prices on the order book. If one were strictly more profitable (in or out of the support), all sellers would deviate to offer only that price. Condition 3 ensures that prices are consistent with agents' expectations about the possible ratios of buyers to sellers.

2.5 Equilibrium solution

We proceed by presenting a candidate equilibrium, then prove that it uniquely satisfies the equilibrium requirements.

As a preliminary, we introduce some notation used throughout the equilibrium solution. First, we define X_t as the probability that a buyer successfully transacts in a given attempt, as derived from the exogenous distribution of buyer/seller ratios:

$$X_t \equiv F(1) + \int_1^\infty \frac{1}{q} F'(q) dq. \tag{4}$$

Note that with probability $F(1)$, all buyers transact, but when $q > 1$, each buyer transacts with probability $\frac{1}{q}$.

When the buyer does transact, her order could be the first to arrive and pay price $a(0)$, or could be the last and pay $a(q)$. As a step in computing the expected price they will pay, we define X_p as follows:

$$X_p \equiv \int_0^1 \left(\int_s^\infty \frac{F'(q)}{q} dq \right) \frac{F(s)}{1 - F(s)} ds \tag{5}$$

This generates a weighted average of the position of the successful buyer s across all realizations q where $q > s$. The weighting $\frac{F(s)}{1-F(s)}$ is relevant for how much the price increases with position s, as we show below.

For the exponential distribution of agents introduced in Section 2.2, these metrics evaluate to $X_t = \phi \log\left(1 + \frac{1}{\phi}\right)$ and $X_p = \frac{1}{2}\left(\log\left((1+\phi)\left(\frac{1+\phi}{\phi}\right)^{\phi^2}\right) - \phi\right)$.

We combine these two metrics in Γ:

$$\Gamma \equiv \frac{1 - F(1)}{\rho + \theta + X_t F(1) + (1 - X_p)(1 - F(1))} \tag{6}$$

Finally, we define Δ as the net gain when a transaction occurs:

$$\Delta \equiv \frac{d_B - d_A}{\rho} - f_t + f_m \tag{7}$$

The first term comes from the transfer of the asset to the buyer who values it more. This is reduced by the net exchange fees.[11]

With this notation in place, we can then provide the equilibrium solution. Prices have the following function of the buyer/seller ratio q:

$$a(q) = \frac{d_A}{\rho} - f_m + \Gamma \cdot \Delta \cdot \left(1 + \frac{\rho + \theta}{1 - F(q)}\right). \tag{8}$$

The probability that a seller transacts when listing price a is:

$$G(a) = \frac{\Gamma \cdot \Delta \cdot (\rho + \theta)}{\left(a - \frac{d_A}{\rho} + f_m\right) - \Gamma \cdot \Delta}. \tag{9}$$

This distribution has a support of $[a(0), a(1)]$, where:

$$a(0) = \frac{d_A}{\rho} - f_m + \Gamma \cdot \Delta \cdot (1 + \rho + \theta) \tag{10}$$

$$a(1) = \frac{d_A}{\rho} - f_m + \Gamma \cdot \Delta \cdot \left(1 + \frac{\rho + \theta}{1 - F(1)}\right). \tag{11}$$

Proposition 1. *Equations 8 through 11 constitute the unique equilibrium in this market.*

The proof is presented in the appendix. We show by construction that the proposal is an equilibrium, and then establish uniqueness by showing that any other price distribution would result in profitable deviations.

2.6 Competition and Limit Order Pricing

The equilibrium solution reveals how the gains from trade are split between buyers and sellers in this order book market. In the absence of trade, a seller would retain the asset and enjoy present utility $\frac{d_A}{\rho}$. When placing a limit order in this market, the seller's expected value of search (as shown in the proof of Proposition 1) is:

$$V_S(a) = \frac{d_A}{\rho} + \Gamma \cdot \Delta. \tag{12}$$

[11]For instance, $f_t - f_m$ would be the exchange revenue. If this is used to exactly cover exchange cost, then the net fees indicate resources lost in facilitating the transaction.

Recall that Δ is the social gain from trade (moving the asset to a person who values it more), net of exchange fees. Thus, Γ indicates (on average) what fraction of those gains accrue to sellers.

In the absence of trade, a buyer would not obtain the asset and have 0 expected utility. By placing a market order, the buyer's expected value of search (as shown in the proof of Proposition 1) is:

$$V_B = (1 - \Gamma)\Delta - \frac{\Gamma \cdot (\rho + \theta)}{1 - F(1)} \cdot \Delta. \tag{13}$$

The first term is complementary to the interpretation of the seller's gains: buyers claim the residual gains from trade. Indeed, if a buyer's utility were only $(1 - \Gamma)\Delta$, one could think of Γ as the endogenously-determined bargaining power of sellers, determining the split of the fixed benefits Δ. In that sense, Γ measures the competitive balance between sellers and buyers, based on their expected relative scarcity.

Additionally, the order book market mechanism requires time to trade, which reduces the benefits relative to frictionless trade, which indicated in the second term of Eq. 13. These frictions disappear if agents are patient and information does not expire, $\rho + \theta = 0$, or if $\Gamma = 0$ so all benefits accrue to the buyers.

With the crucial role of Γ in market outcomes, it is worth exploring what information it encapsulates. First, consider an extreme case where buyers are almost always more abundant than sellers; that is, let $F(1) = 0$ and let $\chi \equiv \int_1^\infty \frac{1}{q} F'(q)dq$. This results in $X_t = \chi$, while $X_p = 0$ (because $F(s) = 0$ for all $s \leq 1$), and thus $\Gamma = \frac{1}{1+\rho+\theta}$. Indeed, in this scenario, sellers extract all surplus from buyers, because of their relative scarcity: $V_S = \frac{d_A}{\rho} + \frac{\Delta}{1+\rho+\theta}$ and $V_B = 0$.

In the other extreme, consider when sellers are almost always more abundant than buyers: that is, let $F(1)$ approach 1. Thus, $X_t = 1$ and any X_p will be finite, so $\Gamma = \frac{0}{\rho+\theta+1+(1-X_p)0} = 0$. Here, buyers extract all the surplus: $V_S = \frac{d_A}{\rho}$ and $V_B = \Delta$. Indeed, the inefficiency is eliminated because buyers never have to postpone their purchases.

For a more general case, we consider the effect of compressing the distribution F around $q = 1$. In the following result, we show that Γ falls when a distribution places more weight (in a first-order stochastic dominance sense) near $q = 1$ (indicating more balance in the arrival of buyers and sellers). We show this in two parts, comparing the effect from the range where $q > 1$ and then the range where $q < 1$.

Proposition 2. *Consider two distributions of participants, F and $\hat{F}$. Let Γ and $\hat{\Gamma}$ be computed under the distribution F and $\hat{F}$, respectively. $\hat{\Gamma}$ is smaller than Γ if:*

1. *$F(q) = \hat{F}(q)$ for all $q \leq 1$, and $F(q) < \hat{F}(q)$ for all $q > 1$, or*

2. *$F(q) > \hat{F}(q)$ for all $q < 1$, and $F(q) = \hat{F}(q)$ for all $q \geq 1$.*

Thus, the competitive balance Γ improves in buyers' favor when there is more balance between buyers and sellers. This is to be expected in the first case, since fewer buyers are left unserved in expectation. Surprisingly, in the second case, buyers are better off when fewer *sellers* are left unserved as well. The higher likelihood to sell induces competition among sellers that reduces their limit order prices.

Of course, Proposition 2 only applies to changes in $F(q)$ that shift weight towards $q = 1$. If the distribution of participants shifts weight towards 1 in some regions and away from 1 in other regions, the impact on the competitive balance Γ would be ambiguous. For example, in our exponential example, an increase in ϕ will increase $F(q)$ at each q, which mixes the first part of the proposition with the opposite of the second part. In that example, the competing effects net out such that Γ is strictly decreasing in ϕ. That is, the first effect dominates, benefiting buyers as fewer of them are left unserved.

2.7 Expected Welfare

We next consider the expected welfare gains in an order book market. First, we establish two extremes—one with no trade and the other with frictionless trade. Autarky would leave the asset with sellers indefinitely, generating a total welfare of $\frac{d_A}{\rho}$ for the seller. At the other extreme, if there were no frictions from the timing of trades, each seller who enters the market would immediately find a buyer, generating a total welfare of $\frac{d_B}{\rho} - f_t - f_m$. The welfare gain from this frictionless trade is the difference between this quantity and the welfare of autarky, which we have already defined as Δ.

We compare this frictionless benchmark to the gains in an order book market. If we always had equal flows of buyers and sellers, the market would generate V_B per buyers and V_S per sellers, compared to 0 and $\frac{d_A}{\rho}$ in autarky, respectively. Thus the welfare gains per transaction are: $V_B + V_S - \frac{d_A}{\rho} = \Delta \cdot \left(1 - \frac{\Gamma \cdot (\rho+\theta)}{1-F(1)}\right)$. However, this neglects potential imbalance in participants: there can only be n_b transactions if there are more sellers than buyers ($n_a > n_b$), and only n_a transactions otherwise. If we rephrase this as welfare per buyer, there will be 1 transaction per buyer if $q < 1$, and $\frac{1}{q}$ transactions per buyer otherwise. We can then average over the possible realizations of q to get expected welfare:

$$W = \Delta \cdot \left(1 - \frac{\Gamma \cdot (\rho + \theta)}{1 - F(1)}\right) \cdot \left(\int_0^1 1 \cdot F'(q)dq + \int_1^\infty \frac{F'(q)}{q}dq\right) \quad (14)$$

$$= \Delta \cdot \left(1 - \frac{\Gamma \cdot (\rho + \theta)}{1 - F(1)}\right) \cdot X_t. \quad (15)$$

Expected welfare has three components. The first, Δ, is the first best outcome from frictionless trade, as previously noted. Note that the other two components are always weakly less than 1. The third term, X_t, is the probability that a buyer transacts. This exclusively reflects the potential for mismatch with too few sellers. The second term (in parenthesis) reduces the gains from trade because of the time delay involved in frictional trade: the impatience of actors or the possible expiration of information can prevent the market from fully realizing the gains from trade because transactions are delayed, possibly beyond the time when trade is useful. However, note that this term goes to 1 if $\rho + \theta \to 0$. The potential mismatch also matters here in Γ and $F(1)$. This parenthetical term would go to $1/(1 + \theta + \rho)$ if $F(1) \to 1$ (which can be seen by substituting in for Γ). For the discussion that follows, we label the frictional component of welfare as Ω:

$$\Omega = 1 - \frac{\Gamma \cdot (\rho + \theta)}{1 - F(1)}. \quad (16)$$

Both Ω and X_t depend on market makers' beliefs about the distribution of the arrival of buyers and sellers. Changes in these beliefs will alter the expected welfare from the operation of the market. We consider how the welfare is affected when the distribution F is compressed around $q = 1$, in parallel to Proposition 2.

Proposition 3. *Consider two distributions of participants, F and $\hat{F}$. Let W and $\hat{W}$ be the expected welfare under the respective distributions. Welfare $\hat{W}$ is higher if:*

1. *$F(q) = \hat{F}(q)$ for all $q \leq 1$, and $F(q) \leq \hat{F}(q)$ for all $q > 1$, or*

2. *$F(q) = \hat{F}(q)$ for all $q \geq 1$, and $F(q) \geq \hat{F}(q)$ for all $q < 1$*

Both cases depict the distribution shifting toward greater parity between buyers and sellers, which increases the number of transactions that occur (all of which have a net welfare gain of Δ). This is true even in the second case, where the $\hat{F}$ distribution makes buyers less scarce in the times that they are the short side of the market. This is a surprising result since welfare is calculated on a per-buyer basis, so having more buyer on average risks diluting the gains; but the additional transactions that occur more than compensate for the larger population.

Of course, Proposition 3 does not consider distributional shifts that alter $F(1)$, the likelihood that all buyers are served. Such a change would need to be evaluated with a particular functional form. For our exponential example as previously noted, an increase in ϕ will shift the competitive balance towards buyers, always reducing Γ. Even so, this can decrease expected welfare because it is more likely that very few sellers match (particularly when $\rho + \theta$ is high). We illustrate this in the next subsection.

2.8 Market liquidity and Welfare

In limit-order markets, an asset is typically defined as being more liquid if it has a smaller spread or if a given execution has a smaller price impact. Often, it is taken for granted that greater liquidity is welfare improving, a conjecture that we examine here.

From Eq. (8), we can calculate the spread and price impact in this market as a function of the model parameters. We define the spread to be

$$S = a(0) - \left(\frac{d_A}{\rho} - f_m\right) = (1 + \rho + \theta) \cdot \Gamma \cdot \Delta \tag{17}$$

which is the difference between the lowest price that the order book ever offers, a_0, and the price that would come from a competitive market.

The price impact for an order that absorbs the fraction q of the order book in our model is defined to be

$$PI(q) = a(q) - a_0 = (\rho + \theta) \cdot \Gamma \cdot \Delta \cdot \frac{F(q)}{1 - F(q)}. \tag{18}$$

Equations (17) and (18) show that market liquidity as characterized by spread and price impact depend on market characteristics as summarized in the following result.

Proposition 4. *Equilibrium spread and price impact decrease when:*

1. *traders become more patient (ρ declines) and/or information is longer-lasting (θ declines), and*

2. *the competitive balance Γ decreases.*

3. *the difference between buyer and seller valuations ($d_B - d_A$, in Δ) decreases,*

4. *the fee spread ($f_m - f_t$ in Δ) decreases,*

In some cases, welfare moves in tandem with market liquidity, but there are notable exceptions, as indicated in the following corollary:

Corollary 1. *In comparative static analysis of equilibrium,*

1. *A decrease in ρ or θ will lead to greater market liquidity (smaller S or $PI(q)$) and greater welfare (W).*

2. *A shift in distribution F to $\hat{F}$, per Propositions 2 and 3, will lead to greater market liquidity and greater welfare.*

3. *A decrease in Δ will lead to greater market liquidity but reduced welfare.*

The first result of Proposition 4 and Corollary 1 indicates that when traders have a lower effective discount rate, the market is more liquid and generates greater welfare. This is because more trades are realized (if θ is lower) or future trades are valued more (if ρ is lower), which amplifies the benefits of trade.

It is more surprising that the division of gains from trade also affect these measures of market liquidity (in claim 2 of Proposition 4). When competitive pressure favors buyers (that is, liquidity takers over market makers), we see smaller spreads and less price impact. Competition among sellers drive them to focus more on the lower prices ($G(a)$ increases) and to reduce their lowest price (a_0 falls).

One example of a distributional change that reduces Γ is when F places shifts weight toward $q = 1$, as established in Proposition 2, which would also increase welfare, per Proposition 3. That is, greater balance in the market not only makes prices more favorable for buyers, but ensures that more transactions are completed due to having similar numbers of participants on both sides.

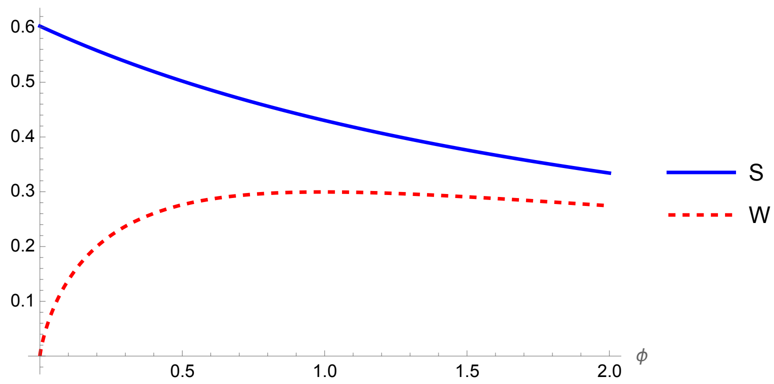


Figure 1: The equilibrium spread and expected welfare as a function of ϕ

The third and fourth result of Proposition 4 indicate that markets are more liquid as Δ falls. The difference in valuations between buyers and sellers can arise through tailored marketing campaigns, substantive differences in the interpretation of available public data, differences in information gathering technologies or preferences amongst market participants, differences in hedging demands due to alternative portfolio goals, or alternative tax or regulatory constraints among market participants. In situations where this heterogeneity is more extreme, we would expect markets to be less liquid.

Additionally, Δ depends on the fee spread $f_m - f_t$ provided to market participants by the exchange. Exchanges where the fee spread is high will be less liquid, *ceteris paribus*.[12] The increased wedge between seller valuations and buyer valuations that arises from increases in the fee spread leads to less liquid markets.

Whether driven by valuation differences or fee spreads, a decrease in Δ will increase market liquidity, but decrease market welfare! Theses perceived differences between buyers and sellers are what motivate trading in the first place, so if the net benefit of a given transaction falls, the expected welfare in aggregate falls proportionally. This suggests that one should be cautious about assuming that market liquidity is a clear signal of market welfare.

This distinction between liquidity and welfare can also occur with more general changes in the distribution of participants, F. In our exponential arrival example, as ϕ increases, it reduces the likelihood of having more buyers than sellers, but this

[12] Colliard and Foucault (2012) build a model that focuses specifically on the efficiency of changes in exchange fees. They find that although smaller fee spreads increase gains from trade, when executions happen, they do not necessarily make markets more efficient.

happens both above and below $q = 1$, so the conditions of Propositions 2 and 3 do not apply. As shown in Figure 1, welfare initially increases with ϕ, but hits a maximum (at $\phi = 1.02$ under these parameters) and then is falling thereafter. In contrast, the equilibrium spread is strictly decreasing as ϕ increases. Thus, when buyers frequently outnumber sellers (ϕ is small), greater liquidity also means greater welfare. But when sellers are more plentiful, the spread (and Γ) continues to fall, but this increased liquidity comes with less welfare.

2.9 Comparison with the Glosten and Milgrom (1985) and Kyle (1985) models

One of the key differences between the model in this paper and the workhorse model of Glosten and Milgrom (1985) and Kyle (1985) is the absence of noise traders in the present model. This absence means that variables like the fraction of traders who are believed to be informed (as in Glosten and Milgrom (1985)) or the relative size of noise trader demand relative to fundamental uncertainty (as in Kyle (1985)) have no analog in our paper. However, the issue of spread size addressed in Glosten and Milgrom (1985) and price impact, addressed in Kyle (1985), can be compared.

Several key pieces of intuition about the formulation of the spread arise from Glosten and Milgrom (1985). These include the idea that competition amongst liquidity providers narrows spreads and that increases in the fraction of traders who are thought to be informed increases spreads. We see a similar result in terms of spread. An increase in the competitiveness of the offer side (as parameterized by an increase in ϕ in the example), leads to narrower spreads. Glosten and Milgrom (1985) also contains the result that increases in the variability of the underlying asset value (as characterized by the difference in the value of the asset in the good state of the world versus the bad state of the world) lead to increases in the spread. While not directly analogous, in our model, differences in the valuation of market makers versus liquidity demanders also lead to increases in the spread and a reduction in market liquidity.

Kyle (1985) characterizes the equilibrium price in the market as an increasing function of the ratio of the underlying asset variation relative to the variance of noise trader demand. If one interprets the difference in values between buyers and sellers in our model as being positively related to the underlying fundamental uncertainty about asset values, then our model and Kyle's model give results that are consistent in that price impact is increasing in this measure. Our model goes further in relating more competition among buyers to a larger price impact.

3 Estimating model parameters with order book data

In this section, we apply the model to order book data from Coinbase's Bitcoin/USD exchange. We first describe key features of the data, then develop a method for extracting analogs of the model parameters from the data. This method includes dividing the data into plausibly steady-state regions called *regimes* that we analyze, first by inspecting four sample regimes and describing the insights that applying the model brings. We then apply the model to each regime and describe the results. In doing so, we decompose the underlying factors driving the observed variation in prices and liquidity across regimes and characterize the drivers of liquidity and welfare in the market.

3.1 Data Description

The Coinbase Bitcoin/USD exchange provides a live feed of order book updates through their Websocket API. These updates provide all changes to the order book in near real-time. In order to demonstrate the technique used here, we limit our analysis to one hour,[13] which yielded 1,321,084 order book updates and 14,807 order executions. These updates occur at the rate of approximately 353 updates per second and 3.95 executions per second.

Our steady-state model is most applicable during periods when the order book is relatively stable for long enough that participants can fully adjust to market conditions. In settings with algorithmic traders, that stability can plausibly be achieved in fractions of a second rather than (as in macro models) months or years. In the context of our model, a steady state order book is not necessarily constant, as the number of available limit orders will fluctuate in any given snapshot due to the inflow of new limit orders or new execution of market orders. Even so, the lowest ask price throughout a steady state period should never fall below the highest bid price over the

[13]Data was collected on June 20, 2023, from approximately 22:25:51.04 UTC to 23:28:18.46 UTC.

Table 1: Regime summary statistics

Regimes ($N = 2480$)	Duration (sec)	Updates (#)
Mean	1.51	533
St. Dev.	3.07	836
Min	0.00	2
25%	0.04	97
50%	0.35	244
75%	1.60	664
Max	35.93	12820

same period; when such a price shift occurs, it sends a clear signal that fundamental parameters have shifted, leading to a new steady state.

We use this idea to characterize the regimes to which we apply our model. Specifically, we begin a regime at the start of trading, and track the highest best bid and the lowest best ask over time. We end the regime and start a new one when either the best bid is higher than the initially (at the start of the regime) observed best ask or the best ask is lower than the initially observed best bid. We repeat this process until the end of the sample. In so doing, we generate 2480 regimes. Some of these regimes are very short, lasting only microseconds, while others last for several seconds.

Table 1 gives statistics on the duration and the number of updates for these regimes. Regime duration varies with the median regime lasting about 0.35 seconds. The longest steady-state regime lasts about 36 seconds and the mean is 1.51 seconds. The median number of order book updates performed during a regime is 244 updates, with a mean of 533 updates and a maximum of 12,820 updates.[14] We anticipate that our model is most applicable to regimes with longer durations; those with extremely short durations occur when both bid and ask prices are rapidly rising or falling, which suggests that the market is still adjusting to changes in fundamental parameters and has not yet reached steady state.

Figure 2 shows the evolution of the best ask (price), the spread, and the price impact (computed for a purchase of 2 BTC) for the steady-state averaged order books, while Table 2 provides summary statistics of the data.

[14]Note that the duration quartiles do not necessarily correspond to the same regime as the corresponding quartile in the distribution of number of updates, since the rate of updates per second is not constant.

Figure 2: Evolution of Price, Spread and Price Impact for Regime-Averaged Order books

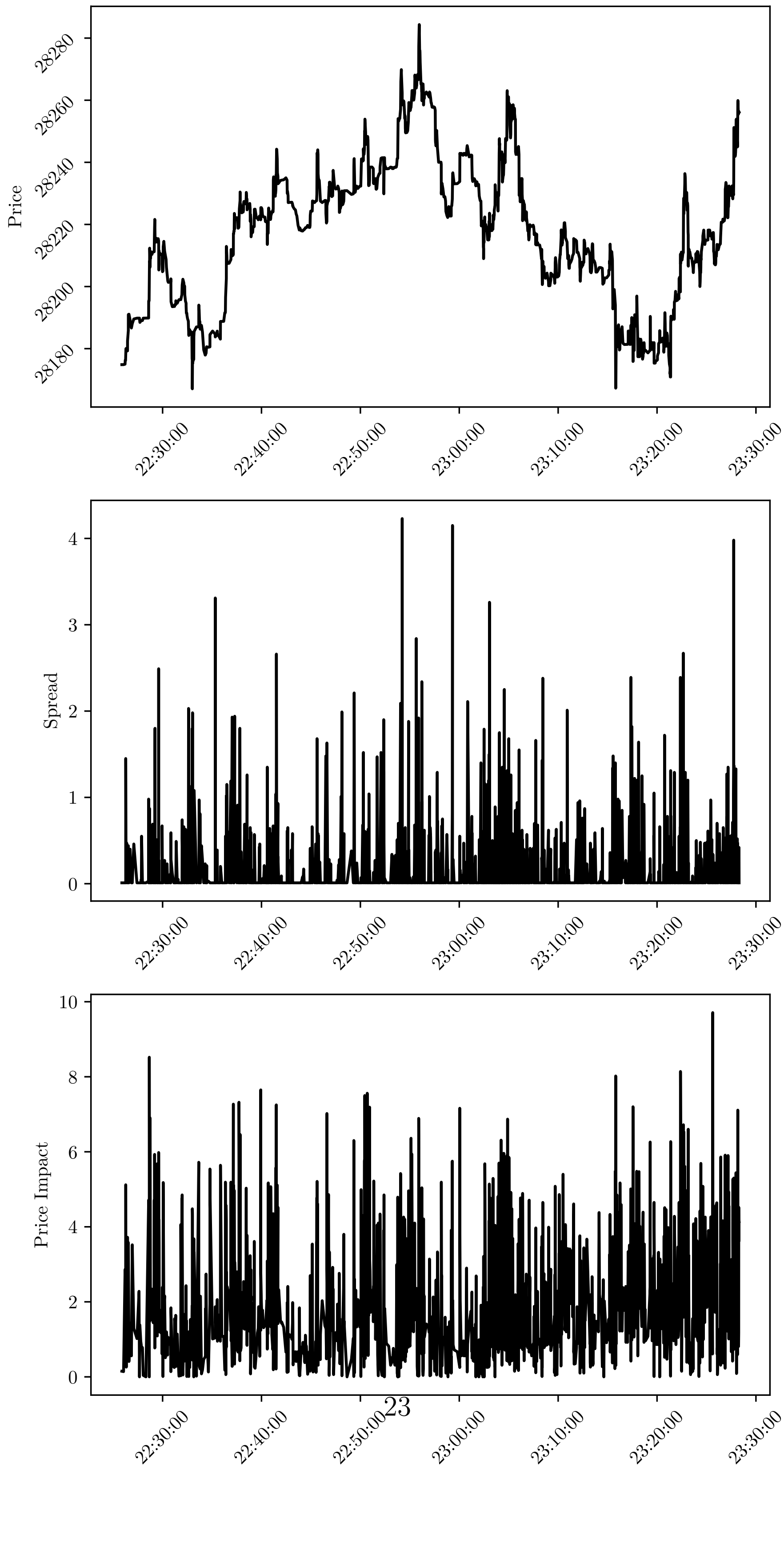


Table 2: Summary statistics of Price, Spread and Price Impact for Averaged Order books

($N = 2480$)	Price	Spread	Price Impact
Mean	\$28,218.92	\$0.16	\$2.12
St. Dev.	\$23.92	\$0.38	\$1.48
Min	\$28,167.07	\$0.01	\$0.00
Median	\$28,219.10	\$0.01	\$1.81
Max	\$28,284.36	\$4.23	\$9.71

During this period, the best available ask price varied from a low of \$28,167.07 to a high of \$28,284.36, which represents a one-hour swing of about 0.38%. The mean and median price were essentially centered in that range. The spread averaged \$0.16, but over half of the regimes had a spread of \$0.01, making the distribution highly skewed. The price impact averaged a \$2.12 increase from a purchase of 2 bitcoins, but this distribution is also left-skewed, with a median price impact of \$1.81. As seen in Figure 2, the price shows a minor trend over the hour, where spread and price impact are more noisy.

3.2 Interpreting data through a steady-state lens

We begin our estimation procedure by calibrating values of ρ and θ. With the rapid pace of transactions in this exchange (with time in units of seconds), we imagine ρ to be a small number, while the average regime duration of 1.5 seconds suggests a calibrated rate of information obsolescence of $\theta = 2/3$. In the calculations that follow we use $\rho + \theta = 0.67$. Next, we turn to the question of what data to consider when estimating the parameters of the order book.

Next, we note that the order book allows for limit orders with arbitrarily large price offers. A market maker can costlessly place these even though they are unlikely to ever be executed. Such dormant limit orders are beyond the scope of our model. Instead, we focus on the portion of the order book that regularly executes. We calibrate $q = 1$ to an execution of 1 BTC (covering 99% of executions, per Table 3), and for calculating statistics of the distribution where $q > 1$, we truncate the distribution at $q = 3$ (e.g. 3 BTC). We later demonstrate that our calculations are robust to these choices.

We can then begin to back out the beliefs of liquidity providers about the distribution $F(q)$. This is separately estimated for each regime, using all observations of the order book throughout that regime to generate an average of the underlying beliefs during the regime.

At any time t, we observe a finite set of limit orders that have been placed on the order book. Let $\mathscr{A}_t$ be an index into this set. Let the set of prices at which there is positive weight on the order book at time t be $\{a_j\}_{j \in \mathscr{A}_t}$ and the set of quantities (normalized by the calibrated value of $q = 1$) be $\{q_j\}_{j \in \mathscr{A}_t}$. We interpret this as a sample from the distribution of sellers. These sellers know that any price on the order book between $a(0)$ and $a(1)$ is equally profitable. Given their beliefs about $F(q)$ and equations (17) and (18), we can back out an implied CDF of q

$$\hat{F}(q) = \frac{\rho + \theta}{(1 + \rho + \theta)\frac{PI(q)}{S} + \rho + \theta}. \tag{19}$$

We observe this implied $\hat{F}(q)$ for each order book in the regime. Using these as observations on the true CDF $F(q)$, we then fit a mixture of normal CDFs to the data by binning the observed qs into n bins where n_b is calculated using Doane's method.[15] Then we divide the data into each bin and select CDFs that have a standard deviation equal to the overall aggregate standard deviation and a mean equal to the mean of samples in that bin. This gives a collection of normal CDFs $\Phi_1, \ldots, \Phi_k$ and the CDF of that regime is taken choosing weights $w_1, \ldots, w_k$ that minimize the sum of squared errors between the observed $F(q)$ and the estimated $\hat{F}(q) = w_1\Phi_1(q) + \cdots + w_k\Phi(q)$ subject to the constraint that $\hat{F}(3) = 1$ and that the CDF be monotonically increasing.[16]

The next step is to use the estimates of $F(q)$ to calculate X_p and X_t per Eqs. 4 and 5, which in turn allow us to compute the competitive balance Γ from Eq. (6).

[15]In Doane's method, the number of bins n_b is given by

$$\begin{aligned} n_b &= 1 + \log_2(N) + \log_2(1 + \frac{|g_1|}{\sigma_{g_1}}) \\ g_1 &= \overline{\left(\frac{x - \mu}{\sigma}\right)^3} \\ \sigma_{g_1} &= \sqrt{\frac{6(N-2)}{(N+1)(N+3)}}. \end{aligned} \tag{20}$$

[16]This method is inspired by Kernel density estimation, but has been adapted because we do not directly observe draws from the density $f(q)$, but instead observed implied values of the CDF $F(q)$.

Table 3: Execution summary statistics

	Execution Size
0.10	0.00025
0.25	0.00177
0.50	0.01777
0.75	0.05318
0.90	0.17716
0.99	0.90437

Finally, we can decompose the spread into the valuation differences between buyers and sellers (Δ) and the measure of seller competitiveness (Γ) using the spread in Eq. (17). Given estimates of $F(q)$, Δ, Γ and X_t, we can calculate the welfare estimate of that regime using equation (14). In this process, optimization for the mixture weights or the numerical process of estimating the market statistics from these distributions does not converge for 14 of the 2480 (0.56%) regimes. We exclude these regimes from the discussion that follows.

3.3 Four Case Studies

To explore more carefully the empirical content of the model, we first look carefully at four specific regimes which provide examples of what our method uncovers about liquidity, welfare, execution probabilities, and estimated beliefs about arrival rates. These regimes were chosen to have similar durations near the average regime duration. Summary statistics for these regimes are given in 4. For each regime, we report the three components of expected welfare: the gap between buyer and seller valuations (Δ), the probability that the buyer can transact in a given period (X_t), and the frictional component stemming from information obsolescence and impatience (Ω, which falls as frictions increase).

First, we compare regimes 1 and 2. These two regimes reflect a period of relative liquidity as indicated by a small spread, but they differ in some important ways. First, while both regimes have a competitive balance Γ that strongly favors buyers, this is much stronger in the first (0.006 vs. 0.018). Second, the welfare per buyer is significantly higher in regime 1 than in regime 2 (\$0.571 vs. \$0.195), but this mostly stems from a larger difference in valuations of the asset (\$0.963 vs. \$0.335). In terms of the fraction of potential gains W/Δ, both regimes are quite similar (59.3%

Table 4: Descriptive statistics of selected regimes

	Seconds	Spread	Δ	Γ	Welfare	X_t	Ω	$\frac{W}{\Delta}$
1	1.509	0.01	0.963	0.006	0.571	0.996	0.595	0.593
2	1.661	0.01	0.335	0.018	0.195	0.987	0.589	0.581
3	1.672	0.25	1.412	0.106	0.733	0.950	0.547	0.519
4	1.641	0.45	1.509	0.179	0.661	0.865	0.507	0.438

and 58.1%) — a consequence of both almost always having buyers transact (X_t) and having similar frictions (Ω). In Figure 3, the solid and dashed lines indicate the estimated arrival beliefs in the two regimes, where regime 2 is more likely to have more buyers arrive (indicated by a lower cumulative distribution at each q). This generates its higher competitive balance Γ, though the difference is fairly small.

Moving to regimes 3 and 4, both have larger spreads with a competitive balance tilting more towards sellers (especially in regime 4). Welfare is also higher than in regimes 1 and 2, but only because the estimated difference in valuations is much larger (\$1.509 and \$1.412, respectively). As a fraction of potential gains, regime 3 only achieves 51.9% while regime 4 achieves only 43.8%. This comes from a lower probability that buyers transact (X_t), as well as worse frictions (lower Ω).[17] In Figure 3, the dot-dashed and dotted lines show that buyers are likely to be more plentiful, tilting the competitive balance toward sellers and making buyers less likely to transact.

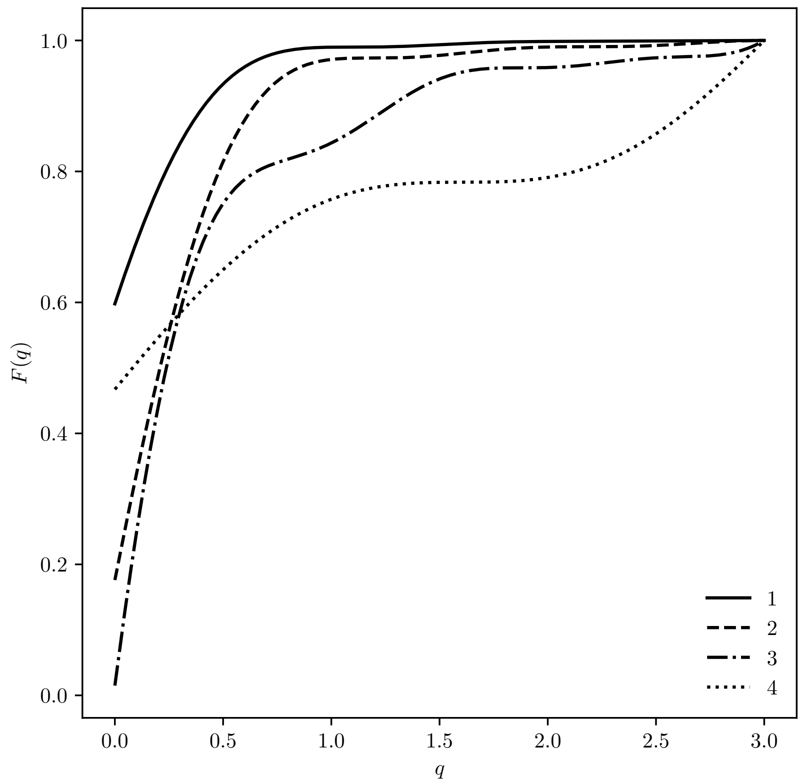
These regime comparisons highlight an important takeaway from this paper, which is that liquidity is not necessarily a proxy for the overall welfare of market participants. Regimes 3 and 4 have the highest welfare per buyer due in large part to their large estimated differences in buyer and seller valuations. In our model, a larger difference in valuations will produce larger spreads and larger welfare (per Propositions 4 and 1), *ceteris paribus*, potentially breaking any correlation between smaller spreads and higher welfare.

3.4 The distributions of Δ, Γ, Welfare and related values

Moving beyond these case studies, we now study the entire sample of estimated regime characteristics. We report the distribution of estimates for valuation differences Δ,

[17]Note that regime 3 does not have the highest value of valuation differences, probability of a buyer transacting, or the friction component; yet the product of these three is largest among the 4 regimes studied.

Figure 3: Estimated distribution of order flows, $F(q)$, for selected regimes



competitive balance Γ, welfare, X_t, the frictional component Ω, and welfare normalized by Δ in Table 5. First, we note that the distribution of valuations differences Δ is highly skewed, with a mean of \$1.51 but a median of \$0.48. This small difference between buyers and sellers seems reasonable, considering the average asset price is \$28,219.

The distribution of Γ shows that market makers were quite competitive throughout this data, with a median of 0.021 and a mean of 0.071. It is interesting to note that for this sample, while Γ has a much lower standard deviation than Δ (0.142 to 2.458), its coefficient of variation is larger. Of the components that make up welfare, X_t and Ω are also less volatile than Δ. As a result, the fraction of potential welfare gains that are actually realized is remarkably stable with a median of 58.0% and a mean of 55.0%. Figure 4 plots the distribution of welfare as well as its components. These

Table 5: Distributions of competitive balance, market welfare, and its components

	Δ	Γ	Welfare	X_t	Ω	W/Δ
Mean	1.511	0.071	0.827	0.962	0.569	0.550
Coeff. of Variation	1.627	2.000	1.642	0.076	0.081	0.135
Min	0.006	0.002	0.001	0.310	0.286	0.167
25%	0.277	0.013	0.156	0.969	0.570	0.552
50%	0.476	0.021	0.273	0.987	0.587	0.580
75%	1.195	0.054	0.632	0.992	0.592	0.587
Max	15.252	0.947	8.818	1.000	0.598	0.597

pictures confirm that estimates of Δ drive the distribution of welfare.

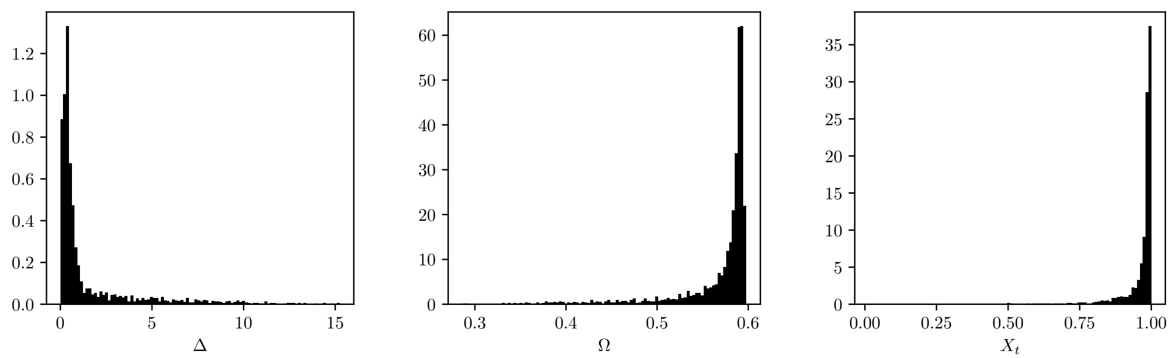
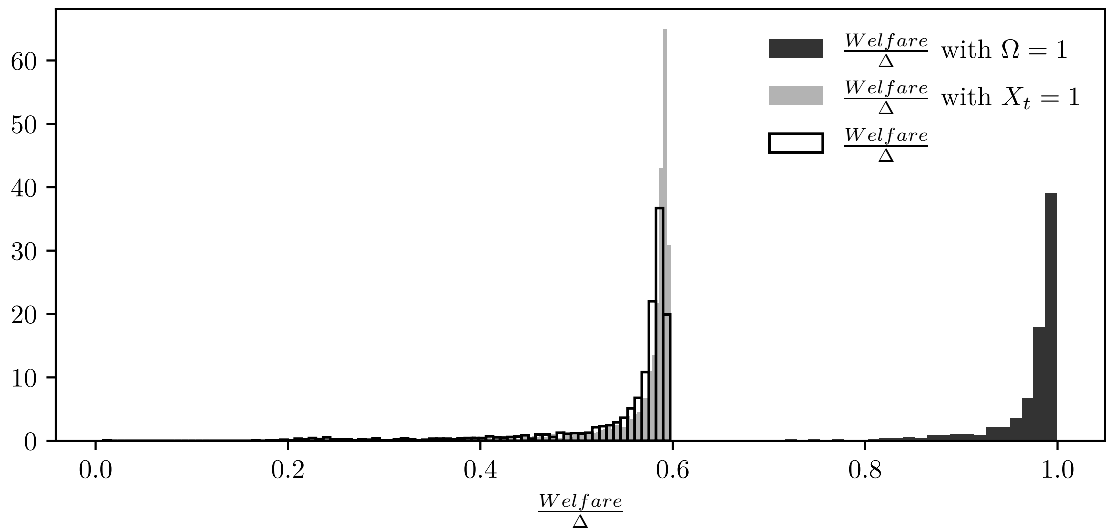


Figure 4: Histograms of the components of welfare

To further assess the components of welfare, we consider two counterfactual experiments. In the first, we consider what would happen if the market did not face any informational obsolescence, eliminating the friction from delayed transactions by setting $\Omega = 1$, but retaining the chance of not matching in a given period. In the second, we retain the frictional component Ω but set $X_t = 1$ for each regime, eliminating the risk of being left unmatched. We then compute the fraction of potential welfare that would be realized in each regime, and compare it to W/Δ estimated from the data. The distribution of estimates across regimes is compared in Figure 5.

Note that the welfare distribution from the data (white) nearly coincides with the estimates under the counterfactual with not risk of buyers being unmatched (gray), indicating that the unmatched buyers contribute only a small amount to welfare losses. This may particularly be true for our sample since the competitive balance skews so heavily towards buyers already, so few are left unmatched in our regimes. In contrast, the frictions from possible obsolescence of information has a major impact; when eliminated, the counterfactual welfare (black) is highly concentrated near 1.

Figure 5: The distribution of $\frac{W}{\Delta}$ (white), and with $\Omega = 1$ (light gray) and $X_t = 1$ (black)



3.5 Understanding the determinants of liquidity

This section explores the determinants of the spread and price impact over the regimes studied here. Equation (17) decomposes the spread into the valuation differences between buyers and sellers (Δ), the measure of seller competitiveness (Γ), and the discounting parameters ρ and θ. Table 6 gives the correlation between the log spread, and logged values of Γ and Δ. Logs are taken because equation (17) suggests that the relationship between these logged values is linear. We see that $\log(\Delta)$ is more highly correlated with spread than Γ. In the estimation process, we use the shape of the order book to estimate Γ and directly observe the spread. Values for Δ are then estimated as a residual. So one lens on the value of the model in understanding the size of the spread is that regressing $\log(S)$ on $\log(\Gamma)$ yields an R^s of 0.284.

To understand the role of competitive balance in the spread, figure 6 shows how the distribution of spread would change if Δ was assumed to stay as it is in the sample and the competitive balance was changed to be $\Gamma = 0.5$ from its estimated values. As can be seen graphically and in Table 7, the mean spread would increase from \$0.160

	log(S)	log(Γ)	log(Δ)
log(S)	1.000	0.533	0.727
log(Γ)	0.533	1.000	-0.193
log(Δ)	0.727	-0.193	1.000

Table 6: Correlation between determinants of liquidity

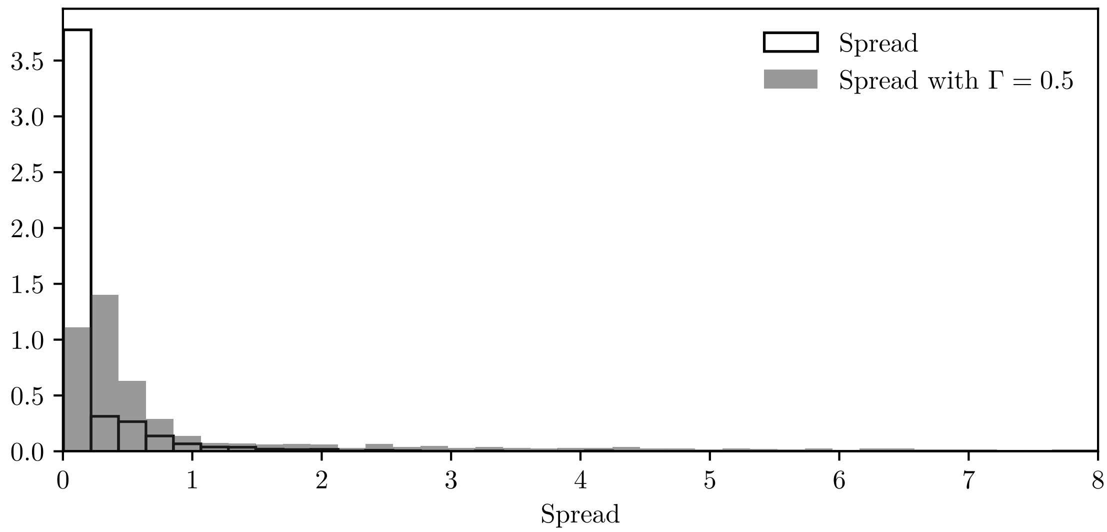


Figure 6: The predicted distribution of spread if competitive balance were $\Gamma = 0.5$

to \$1.262, and the standard deviation would increase from \$0.374 to \$2.053. The median spread in this world would be nearly 40 times higher, going from \$0.01 to \$0.397. This suggests that while estimated valuation differences are very important in understanding the spread over this sample, the competitive balance component of the spread is also an important phenomenon to understand.

3.6 Robustness

We focus our discussion of robustness on the impact on welfare estimates of alternative parameterizations of the number of units of trade that correspond to $q = 1$, and the truncation of the distribution corresponding to the place where $F(q) = 1$. In the results section we chose 1 BTC to be the value at which $q = 1$ and $q_{max} = 3$ to be the truncation point. Figure 7 shows the distribution of welfare and $\frac{W}{\Delta}$ for $q \in \{1, 1.5, 2\}$, while Table 8 gives summary statistics for these distributions. The mean welfare is

	Spread	Spread ($\Gamma = 0.5$)
Mean	0.160	1.262
St. Dev.	0.374	2.053
min	0.010	0.005
25%	0.010	0.232
50%	0.010	0.397
75%	0.080	0.998
Max	4.230	12.735

Table 7: The predicted distribution of spread if competitive balance were $\Gamma = 0.5$

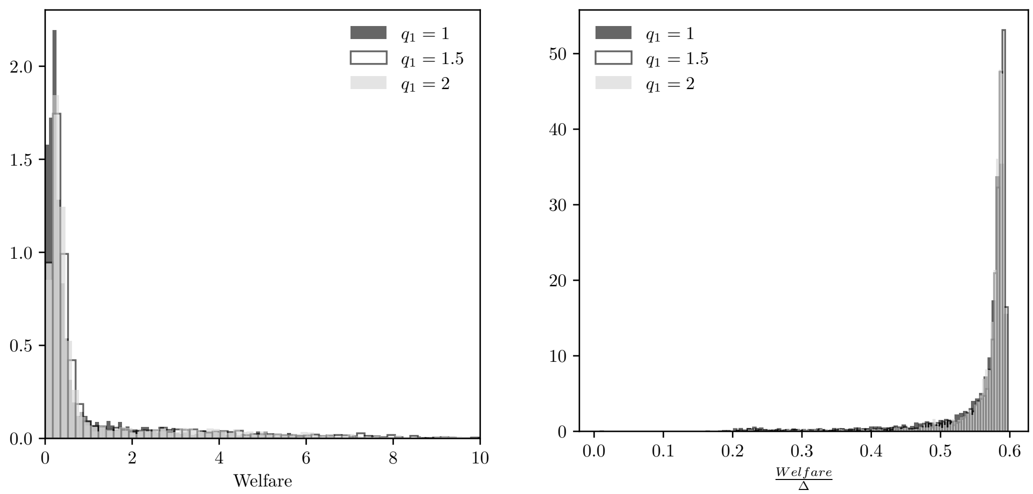


Figure 7: The distribution of welfare (left) and $\frac{W}{\Delta}$ (right) for $q = 1$ corresponding to BTC values in $\{1, 1.5, 2\}$

higher for larger BTC values associated with $q = 1$. This happens because as the value associated with $q = 1$ increases, buyers are much more likely to be interpreted as being able to successfully buy. While the means vary across parameterizations, the median value of potential welfare gains that are realized is similar across the three values of $q = 1$ and the the median value of W/Δ is even closer. This suggests that the results of this model are relatively robust to alternative paramterizations of the value at which $q = 1$, and that these differences are driven by some of the extreme welfare values.

Figure 8 shows the distribution of welfare and $\frac{W}{\Delta}$ for $q_{max} \in \{2, 3, 4\}$ with the value BTC values corresponding to $q = 1$ as either 1, 1.5 or 2. Table 9 gives summary statistics for these distributions. The mean welfare decreases for larger values of q_{max}

	Welfare			$\frac{W}{\Delta}$		
	$q_1 = 1$	$q_1 = 1.5$	$q_1 = 2$	$q_1 = 1$	$q_1 = 1.5$	$q_1 = 2$
Mean	0.826	1.201	1.288	0.550	0.565	0.569
St. Dev.	1.358	1.981	2.086	0.074	0.055	0.045
Min	0.000	0.001	0.001	0.008	0.202	0.198
25%	0.155	0.224	0.227	0.552	0.570	0.570
50%	0.273	0.372	0.366	0.580	0.584	0.583
75%	0.631	0.887	0.973	0.587	0.590	0.588
Max	8.818	17.482	15.442	0.597	0.597	0.596

Table 8: The distribution of welfare (left) and $\frac{W}{\Delta}$ for $q = 1$ corresponding to BTC values in $\{1, 1.5, 2\}$

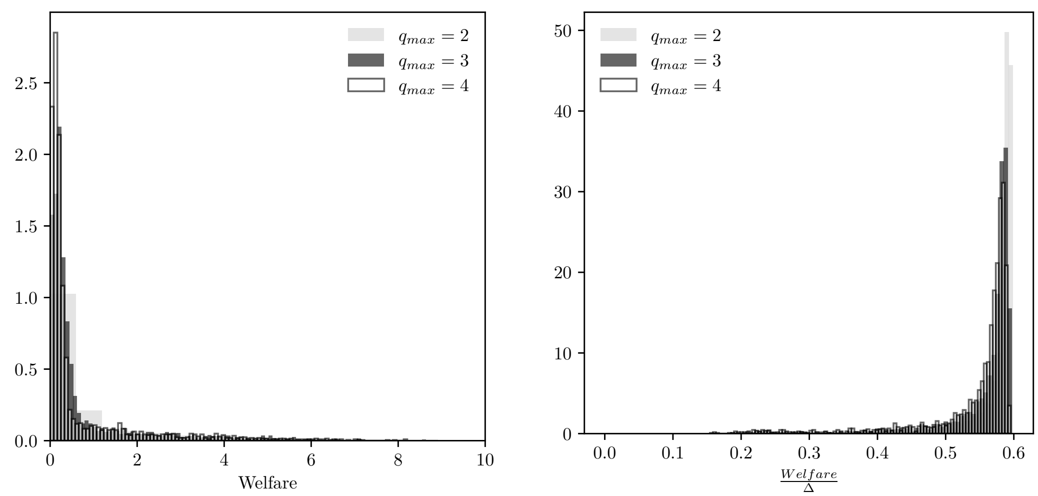


Figure 8: The distribution of welfare (left) and $\frac{W}{\Delta}$ (right) for $q_{max} \in \{2, 3, 4\}$

as increasing q_{max} tends to decrease the fitted value of $F(1)$, making it appear less likely that sellers will be able to sell. The median value of potential welfare gains changes more than changes in the value corresponding to $q = 1$, although these results are not qualitatively different. The median value of W/Δ is quite stable across these parameterizations.

Taken together, these robustness results suggest a model that is qualitatively stable to changes in the parameterizations considered here.

	Welfare			$\frac{W}{\Delta}$		
	$q_{max} = 2$	$q_{max} = 3$	$q_{max} = 4$	$q_{max} = 2$	$q_{max} = 3$	$q_{max} = 4$
mean	1.791	0.826	0.690	0.572	0.550	0.544
std	4.527	1.358	1.224	0.048	0.074	0.073
min	0.000	0.000	0.001	0.000	0.008	0.154
25%	0.221	0.155	0.101	0.575	0.552	0.543
50%	0.414	0.273	0.196	0.589	0.580	0.572
75%	1.373	0.631	0.474	0.593	0.587	0.582
max	59.418	8.818	8.209	0.599	0.597	0.596

Table 9: The distribution of welfare (left) and $\frac{W}{\Delta}$ for $q_{max} \in \{2, 3, 4\}$

4 Conclusion

Order book markets create different roles for liquidity providers and demanders, where the former set a price and await a transaction, while the latter show up and take the best available price. Our model is built on this essential feature. Both parties must anticipate the arrival rate of market participants to predict the wait length at a given price (for sellers) or the likely price (for buyers). Thus, the observed order book (including spread and price impact) are an equilibrium consequence of beliefs about arrival rates. This distribution of arrival rates can be summarized in the competitive balance, which determines how the gains from trade are split between buyers and sellers. We demonstrate that as the competitive balance favors sellers, the market liquidity (by traditional measures) falls. Even so, less liquidity can generate greater welfare when it means more parity in the flow of buyers and sellers, leading to a greater volume of trades.

Using data from the Coinbase BTC/USD exchange, we demonstrate how order book data can be interpreted in the light of our model. We identify time periods (regimes) during which the market plausibly appears to be in steady state. As a key check of the model, we find that realized transactions are less frequent when the model-inferred arrival of buyers is low, even though the latter only is derived from price data rather than participant or transaction flows. We estimate that market currently produces a welfare that is 83% of what a frictionless market could generate (on average). Across the regimes in our data, we find that the variation in welfare is largely driven by variation in valuation differences, but that other frictions in the market also play an important role in reducing welfare.

Further work in this area could investigate the competitive balance across alternative exchanges of the same asset. This work would allow us to characterize the pricing and welfare efficiency across exchanges. Such an effort could yield insights into how exchanges compete, the welfare gains and/or losses associated with this competition, and the role of market regulation in promoting such competition. This model also allows for the study of changes in fee structures or other competitive factors across regimes and their potential welfare and liquidity effects.

A Proof of Proposition 1

Proof. First, we verify that the proposal satisfies the equilibrium conditions.

Note that by substituting Eq. 8 into Eq. 9, we readily obtain $G(a(q)) = 1 - F(q)$, satisfying condition 3.

Moreover, since $F'(q) > 0$ for $q \in [0, 1]$, the support of G will run from $a(0)$ (as F is not defined for $q < 0$) to $a(1)$ (as G is not defined for $q > 1$), fulfilling condition 1.

To verify condition 4, we substitute $a(s)$ from Eq. 8 into Eq. 3:

$$\begin{aligned}(\rho+\theta)V_B &= \int_0^1 \left[\int_0^q \frac{1}{q}\left(\Delta - \Gamma\cdot\Delta\left(1+\rho+\theta+\frac{(\rho+\theta)F(s)}{1-F(s)}\right) - V_B\right) ds\right] F'(q)dq \\ &+ \int_1^\infty \left(\int_0^1 \left(\Delta - \Gamma\cdot\Delta\left(1+\rho+\theta+\frac{(\rho+\theta)F(s)}{1-F(s)}\right) - V_B\right) ds\right) \frac{1}{q}F'(q)dq.\end{aligned}$$

After moving out terms that are constant w.r.t. integration, this simplifies to:

$$\begin{aligned}(\rho+\theta)V_B &= (\Delta - \Gamma\cdot\Delta\,(1+\rho+\theta) - V_B)\left(F(1) + \int_1^\infty \frac{F'(q)}{q}dq\right) - \Gamma\Delta(\rho+\theta)\cdot \\ &\left(\int_0^1 \left(\int_0^q \frac{F(s)}{q(1-F(s))}ds\right) F'(q)dq + \int_1^\infty \left(\int_0^1 \frac{F(s)}{q(1-F(s))}ds\right) F'(q)dq\right).\end{aligned}$$

Changing the order of integration allows the last two double integrals to be combined:

$$\begin{aligned}(\rho+\theta)V_B &= (\Delta - \Gamma\cdot\Delta\,(1+\rho+\theta) - V_B)\left(\int_0^1 F'(q)dq + \int_1^\infty \frac{F'(q)}{q}dq\right) \\ &-\Gamma\Delta(\rho+\theta)\cdot\left(\int_0^1 \left(\int_s^\infty \frac{F(s)}{q(1-F(s))}F'(q)dq\right) ds\right).\end{aligned}$$

We then utilize the notation for X_t and X_p:

$$(\rho+\theta+X_t)V_B = \Delta\left(\left(1-\Gamma\left(1+\rho+\theta\right)\right)X_t - \Gamma(\rho+\theta)X_p\right). \tag{21}$$

Since the reservation price makes buyers indifferent about continued search, $\bar{a} = \frac{d_B}{\rho} - f_t - V_B$. At the same time, this must be the highest price offered in equilibrium: $\bar{a} = \frac{d_A}{\rho} - f_m + \Gamma\cdot\Delta\cdot\left(1+\frac{\rho+\theta}{1-F(1)}\right)$. These combine to yield:

$$V_B = \Delta\left(1 - \Gamma\left(1+\frac{\rho+\theta}{1-F(1)}\right)\right). \tag{22}$$

Substituting this into Eq. 21 and rearranging, we obtain:

$$\Delta(\rho+\theta)\cdot(1-F(1)-\Gamma\left(\rho+\theta+F(1)X_t+(1-X_p)(1-F(1))\right))=0$$

If this is solved for Γ, we obtain its definition from Eq. 6. Thus, the maximum price $a(1)$ is exactly the optimal reservation price $\bar{a}$ given the equilibrium price distribution.

Finally, for condition 1, consider any price a in the support of $G(a)$. The seller's expected profit from Eq. 1 rearranges as:

$$V_S(a)=\frac{d_A}{\rho}+\frac{G(a)\left(\rho(a+f_m)-d_A\right)}{\rho(\rho+\theta+G(a))}. \tag{23}$$

Substitute for $G(a)$ from Eq. 9 yields:

$$V_S(a) \;=\; \frac{d_A}{\rho}+\frac{(\rho+\theta)\cdot\Gamma\cdot\Delta\cdot(\rho(a+f_m)-d_A)}{\rho\left((\rho+\theta)\left(a-\frac{d_A}{\rho}+f_m-\Gamma\cdot\Delta\right)+\Gamma\cdot\Delta\cdot(\rho+\theta)\right)}=\frac{d_A}{\rho}+\Gamma\cdot\Delta.$$

This is constant w.r.t. a, so all prices in the support are equally profitable.

If a firm offered a price below $a(0)$, they would transact with the same probability (that is, with probability 1) as when offering $a(0)$, but with a strictly lower price and thus less profit. If a firm offered a price above $a(1)$, it will always be rejected since it is above the reservation price. Thus, everything in the support is profit maximizing and nothing outside of it is.

To see that this is the unique equilibrium, we first prove that there cannot be an equilibrium solution that contains atoms. Suppose at price $\hat{a}\geq a(0)$ there is an atom of weight $g(\hat{a})>0$. A seller who places a limit order of $\hat{a}$ will transact with probability $G(\hat{a})$, after all other sellers with orders at or below $\hat{a}$, earning expected profit:

$$E\pi(\hat{a})=G(\hat{a})\left(\hat{a}+f_m-V_S\right). \tag{24}$$

If the seller instead places the limit order at $\hat{a}-\epsilon$ for some $\epsilon>0$, then the probability of transacting would be $G(\hat{a}-\epsilon)$ and the expected profit would be

$$E\pi(\hat{a}-\epsilon)=G(\hat{a}-\epsilon)\left(\hat{a}-\epsilon+f_m-V_S\right). \tag{25}$$

Note that $\lim_{\epsilon\searrow 0}G(\hat{a}-\epsilon)=G(\hat{a})+g(\hat{a})$, so that the seller is jumping ahead of the mass of sellers with orders at $\hat{a}$. Thus, there exists a sufficiently small $\epsilon>0$ for which

$E\pi(\hat{a}-\epsilon) > E\pi(\hat{a})$. Thus, the expected profit from placing an order at $\hat{a}-\epsilon$ is strictly greater than the expected profit from placing an order at $\hat{a}$ which violates the equal profit condition and hence is not an equilibrium.

For atom-less distributions over $G(a)$, the interior of the support is unique because of the equal profit condition for sellers. The equal profit condition when applied to Eq. (23) implies a differential equation for $G(a)$,

$$\frac{(\theta+\rho)G'(a)\left(\rho\left(a+f_m\right)-d_A\right)}{\rho(G(a)+\theta+\rho)}+G(a)=0. \tag{26}$$

This equation has a unique solution.[18] Thus, $G(a)$ is uniquely defined. Then, the equilibrium requirement that $G(a)$ be consistent with the distribution $F(q)$ given in Eq. (2) uniquely defines $a(q)$. In this relationship, since $F(q)$ is assumed to be strictly monotonic and $G(a)$ is shown to be also monotonic, $a(q)$ is unique.

Last, we note that the endpoints of the support of a are uniquely defined. That $a(1)$ is unique follows from the discussion following Eq. (21). That $a(0)$ has a unique solution follows from solving Eq. (9) using the fact that $G(a(0)) = 1$, which follows from the assumption that $F'(0) > 0$. □

[18]The general solution for a first-order differential equation of this form is

$$G(a)=-\frac{(\theta+\rho)e^{c_1(\theta+\rho)}}{-\rho\left(a+f_m\right)+d_A+e^{c_1(\theta+\rho)}}, \tag{27}$$

while the constant c_1 is pinned down by the necessity that $G(a(0)) = 1$.

B Proof of Proposition 2

Proof. **Part 1**: With rearrangement of the components of X_t and X_p, we get:

$$\begin{aligned}
\Gamma &= \frac{1}{\frac{\rho+\theta}{1-F(1)} + \frac{X_t F(1)}{1-F(1)} + 1 - X_p} \\
&= 1/\left(1 + \frac{\rho+\theta}{1-F(1)} + \frac{(F(1)+\chi)F(1)}{1-F(1)} - \int_0^1 \left(\int_s^1 \frac{F'(q)}{q}dq + \chi\right)\frac{F(s)}{1-F(s)}ds\right) \\
&= 1/\left(1 + \frac{\rho+\theta}{1-F(1)} + \frac{(F(1)+\chi)F(1)}{1-F(1)} - \int_0^1 \left(\int_s^1 \frac{F'(q)}{q}dq\right)\frac{F(s)}{1-F(s)}ds\right. \\
&\qquad \left. -\chi\int_0^1 \frac{F(s)}{1-F(s)}ds\right) \\
&= 1/\left(1 + \frac{\rho+\theta}{1-F(1)} - \frac{F(1)^2}{1-F(1)} - \int_0^1 \frac{F'(q)}{q}\left(\int_0^q \frac{F(s)}{1-F(s)}ds\right)dq\right. \\
&\qquad \left. +\chi\left(\frac{F(1)}{1-F(1)} - \int_0^1 \frac{F(s)}{1-F(s)}ds\right)\right)
\end{aligned}$$

Note that $\chi \equiv \int_1^\infty \frac{1}{q}F'(q)dq$ is larger under $\hat{F}$ as it places more weight where $1/q$ is bigger. All other terms will be the same under F or $\hat{F}$ since they only involve $q \leq 1$. The last parenthetical element is positive because $F(1) > F(s)$ for all $s < 1$, so $\frac{F(1)}{1-F(1)} > \frac{F(s)}{1-F(s)}$. Thus, the denominator of this rearranged expression is larger under $\hat{F}$, and hence $\hat{\Gamma} < \Gamma$.

Part 2: Using the same rearrangement, in the second line, χ is the same under F or $\hat{F}$. Moreover, $\frac{F(s)}{1-F(s)} > \frac{\hat{F}(s)}{1-\hat{F}(s)}$, so the second parenthetical term is larger under $\hat{F}$. In the first line, the second integral is smaller for the same reason for each q, and furthermore $\hat{F}'$ will place more weight where there is a smaller $1/q$. Thus the double integral is smaller, and with the negative, the whole term is larger under $\hat{F}$. Thus, the denominator of this rearranged expression is larger under $\hat{F}$, and hence $\hat{\Gamma} < \Gamma$. □

C Proof of Proposition 3

Proof. By applying Proposition 2, we find that $\hat{\Gamma}$ is smaller than Γ in either case, while $\frac{\rho+\theta}{1-F(1)}$ is unaffected, so the second component of welfare increases. Meanwhile, X_t increases under the first scenario, as $\hat{F}(q)$ places more weight where $\frac{1}{q}$ is larger.

X_t is unchanged in the second scenario since $F(1) = \hat{F}(1)$. □

References

BACK, KERRY AND SHMUEL BARUCH (2004): "Information in securities markets: Kyle meets Glosten and Milgrom," *Econometrica*, 72 (2), 433–465.

BUTLER, ALEXANDER W, GUSTAVO GRULLON, AND JAMES P WESTON (2005): "Stock market liquidity and the cost of issuing equity," *Journal of Financial and Quantitative Analysis*, 40 (2), 331–348.

COLLIARD, JEAN-EDOUARD AND THIERRY FOUCAULT (2012): "Trading Fees and Efficiency in Limit Order Markets," *Review of Financial Studies*, 25 (11), 3389–3421.

CUI, WEI AND SÖREN RADDE (2016): "Money and asset liquidity in frictional capital markets," *American Economic Review*, 106 (5), 496–502.

DUGAST, JÉRÔME (2018): "Unscheduled news and market dynamics," *The Journal of Finance*, 73 (6), 2537–2586.

EASLEY, DAVID AND MAUREEN O'HARA (1987): "Price, trade size, and information in securities markets," *Journal of Financial Economics*, 19 (1), 69–90.

FOUCAULT, THIERRY, OHAD KADAN, AND EUGENE KANDEL (2005): "Limit order book as a market for liquidity," *Review of Financial Studies*, 18 (4), 1171–1217.

GLOSTEN, LAWRENCE R AND PAUL R MILGROM (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14 (1), 71–100.

GOETTLER, RONALD L, CHRISTINE A PARLOUR, AND UDAY RAJAN (2005): "Equilibrium in a dynamic limit order market," *Journal of Finance*, 60 (5), 2149–2192.

HOLLIFIELD, BURTON, ROBERT A MILLER, AND PATRIK SANDÅS (2004): "Empirical analysis of limit order markets," *The Review of Economic Studies*, 71 (4), 1027–1063.

Hollifield, Burton, Robert A Miller, Patrik Sandås, and Joshua Slive (2006): "Estimating the gains from trade in limit-order markets," *The Journal of Finance*, 61 (6), 2753–2804.

Kerr, Jon, Gil Sadka, and Ronnie Sadka (2020): "Illiquidity and price informativeness," *Management Science*, 66 (1), 334–351.

Kyle, Albert S (1985): "Continuous auctions and insider trading," *Econometrica*, 1315–1335.

Lagos, Ricardo and Guillaume Rocheteau (2009): "Liquidity in asset markets with search frictions," *Econometrica*, 77 (2), 403–426.

Pástor, L'uboš and Robert F Stambaugh (2003): "Liquidity risk and expected stock returns," *Journal of Political Economy*, 111 (3), 642–685.

Roşu, Ioanid (2009): "A dynamic model of the limit order book," *Review of Financial Studies*, 22 (11), 4601–4641.

Vayanos, Dimitri and Tan Wang (2007): "Search and endogenous concentration of liquidity in asset markets," *Journal of Economic Theory*, 136 (1), 66–104.

Weill, Pierre-Olivier (2020): "The Search Theory of Over-the-Counter Markets," *Annual Review of Economics*, 12, 747–773.