Information Transparency and Market Efficiency in Blockchain-enabled Marketplaces: Role of Traders' Analytical Ability

Hong Zhang
The University of Texas at Dallas, hong.zhang@utdallas.edu

Zhiqiang Zheng
University of Texas, ericz@utdallas.edu

Amit Mehra
University of Texas at Dallas, amit.mehra@utdallas.edu

Information Transparency and Market Efficiency in Blockchain-enabled Marketplaces: Role of Traders' Analytical Ability

Completed Research Paper

Hong Zhang
Jindal School of Management
University of Texas at Dallas
800 W Campbell Rd, Richardson, TX
hong.zhang@utdallas.edu

Zhiqiang Zheng
Jindal School of Management
University of Texas at Dallas
800 W Campbell Rd, Richardson, TX
ericz@utdallas.edu

Amit Mehra
Jindal School of Management
University of Texas at Dallas
800 W Campbell Rd, Richardson, TX
amit.mehra@utdallas.edu

Abstract

Classic economic theory asserts that full information transparency entails information symmetry and, thus, market efficiency. We test if this theory still holds in a blockchain-enabled marketplace where full information transparency is accomplished. We leverage the data from EnjinX, a non-fungible-token (NFT) marketplace, where the entire historical NFT transactions are symmetrically accessible to all buyers and sellers. We surprisingly observe substantial market inefficiencies. To explain this paradox that inefficiencies persist even in a fully information-transparent environment, we propose that traders' limited analytical ability, rather than information asymmetry, ultimately drives market inefficiencies. We quantify analytical ability by examining whether traders' performance can be augmented by machine-learning algorithms. And we find that having ten more historical transactions increases market efficiency by 1.10%. However, market efficiency could decrease by 69.02% when traders cannot effectively consume the available information. Our findings contribute to the literature by quantifying analytical ability and highlighting the analytical-ability divide phenomenon.

Keywords: Information transparency, analytical-ability divide, market efficiency, blockchain-enabled marketplaces, machine learning augmentation

Introduction

Information asymmetry, the imbalance of information possessed by buyers and sellers, poses a constant challenge to two-sided markets, leading to asset mispricing, market inefficiency, or even market failure (Akerlof 1970; Huang et al. 2022). According to the Efficient Market Hypothesis (EMH), a perfect market, in which everyone receives a free and instant transmission of true and complete historical market information, will create equal opportunities for buyers and sellers to trade, thus rendering an efficient market (Fama 1970). An efficient market could improve social welfare and eventually benefit the whole society. Thus, improving "market efficiency" has become a common goal for platforms and policymakers (Subramanian and Overby 2017).

This paper focuses on studying the market efficiency of blockchain-enabled marketplaces. These marketplaces function similarly to traditional two-sided e-commerce platforms, facilitating asset trading and ownership transferring. However, the use of blockchain cultivates an unparalleled trading environment characterized by the full transparency of market trading information. This transparency in trading is manifested in two ways. First, the full trading history of *each trader account* and *each asset* is ***completely*** transparent to the public. All tick-by-tick level transaction data (i.e., who bought/sold what assets, at what time, and at what prices) are recorded and disclosed on the blockchain. Second, the same market information is distributed *equally* to all buyers and sellers at the same time at no additional costs, thus creating a fully *information-symmetric* environment. That is, the trading information recorded on the blockchain is equally accessible to all market participants. These two unique features regarding trading information transparency, namely *completeness and symmetry*, make blockchain-based marketplaces distinctive from traditional trading marketplaces (e.g., the NYSE stock market and Amazon.com), in that traders could be perfectly and uniformly informed about market trading conditions (such as supply/demand dynamics) and information asymmetry in trading is therefore eliminated in theory.[1]

According to EMH, we would expect these blockchain-enabled marketplaces to be fully efficient due to the elimination of information asymmetry by blockchain. We first investigate if this is true. Toward this end, we examine a specific blockchain-enabled marketplace, EnjinX JumpNet, a leading non-fungible token (NFT) marketplace for trading digital gaming assets in the Enjin Metaverse.

Paradoxically, we still observe excessive market inefficiencies in the EnjinX marketplace using its NFT transaction data from 2021 to 2022. This observation necessitates a re-investigation of EMH; it reflects a basic paradox that inefficiencies may persist even in information-transparent and symmetric markets, thus challenging the conventional EMH wisdom. Understanding why this is happening and where the inefficiencies come from can fundamentally contribute to the information economics literature.

In this paper, we identify two potential inefficiency sources to explain this paradox: (1) the amount of transaction information on blockchain and (2) the level of traders' analytical ability (i.e., the ability to analyze the available blockchain information). The latter is particularly note-worthy for blockchain-enabled marketplaces because these nascent markets often lack a mature infrastructure to support traders in analyzing trading data, compared to traditional established financial markets, such as the stock market. For example, traditional trading markets typically benefit from having a mature financial analysis system, including financial analysts' forecasts and tools in the Institutional Brokers' Estimate System (IBES), to guide investors in trading. In contrast, blockchain-based markets generally leave traders to their own discretion, letting traders rely on their own analytical prowess to decipher historical trading information.

Further, we are also interested in the interaction between these two inefficiency sources, because the effect of transaction information itself on market efficiency also hinges on traders' heterogeneous analytical ability to leverage it. We thus examine three research questions:

RQ1 (Information Transparency Effect): How does the amount of information available on blockchain influence market efficiency?

RQ2 (Trader Analytical Ability Effect): How does traders' analytical ability impact market efficiency?

RQ3 (Interaction Effect): Whether and how does traders' analytical ability moderate the effect of information amount on market efficiency?

To answer the three questions, we measure the two sources of inefficiency as follows. First, we operationalize the amount of blockchain transaction information as the cumulative number of historical sale transactions recorded on the blockchain. Second, for traders' analytical ability, we propose a novel counterfactual-based measure to quantify it using the *contraction* method, which was originally proposed by Kleinberg et al. (2018) and further applied by Fu et al. (2021) and Shen et al. (2022) in the contexts of crowd lending and social trading. The gist of this method is to employ a machine learning (ML) algorithm

[1] Here, we mean the transparency and symmetry of market trading information facilitated by blockchain. We acknowledge that certain external information (e.g., asset consumption in the EnjinX games) might not be stored on the blockchain. Accessing and verifying the authenticity of such off-chain data is commonly known as the "oracle problem."

to augment traders' decisions by removing their suboptimal trades and then compare the performance of ML-augmented decisions with that of decisions made by pure human traders. The performance of the ML-augmented decisions is then used as the counterfactual benchmark to measure traders' analytical ability. Those traders whose performance cannot be further improved by the state-of-the-art ML are considered to have high analytical abilities.

Our analysis reveals that an increase in the amount of historical trading information by ten units improves market efficiency by 1.10%. However, market efficiency may decrease by 69.02% if traders lack the analytical ability to fully utilize the rich blockchain information, despite its transparency to all. Moreover, the presence of low-ability traders may further reduce the impact of blockchain transaction information on market efficiency by 0.11%, indicating that analytical ability positively moderates the impact of information quantity. We further employed various identification strategies to address endogeneity problems and found these results to be robust. Overall, these three findings establish our story that high information transparency and information symmetry do not necessarily warrant high market efficiency. Market efficiency also critically hinges on who the traders are, especially their analytical ability to leverage the transparency of historical information for future trading.

Notably, our paper is among the first to *quantify* "analytical ability" rigorously using an innovative ML counterfactual framework, which has never been documented in the previous literature. This approach allows for a more reliable measurement of analytical ability by creating a quasi-random design where ML-augmented decisions (the treated group) are compared against human decisions (the control group), effectively controlling for the impacts of various confounding factors.

This paper also presents a novel contribution to the extant literature on the digital-divide phenomenon (e.g., Burtch and Chan (2019); Wei et al. (2011)). While previous studies have focused on unequal access to digital information and technologies and their impact on individual productivity and performance, our study sheds light on a new type of divide: the analytical-ability divide, echoing the notion of the "capability divide" proposed in Wei et al. (2011). Further, this paper empirically examines the impact of the analytical-ability divide on market efficiency, a market-level outcome that is rarely explored in the digital-divide literature. Our findings expand the understanding of the digital divide by suggesting that the analytical-ability divide is a significant and new hurdle to market efficiency in the emerging blockchain era and in order to close this divide, platforms are expected to provide sophisticated analytical tools to help traders make decisions.

Literature Review and Hypothesis Development

In this section, we theorize each hypothesis and review relevant literature.

Market Efficiency and Flipping Opportunity

In this paper, we define an efficient market as one in which any flipping opportunities could be promptly exploited away for profit (Malkiel 2003; Subramanian and Overby 2017). Flipping, a term commonly used in the real-estate domain, refers to buying an asset and reselling it expeditiously to make a profit (Bayer et al. 2020). When there is a considerable price difference between two sales of an instrument within a short period, a flipping opportunity arises, which traders (flippers or arbitrageurs) could have exploited for a profit (Overby and Clarke 2012; Subramanian and Overby 2017). However, the "law of one price" dictates that flipping opportunities cannot last long in an efficient market, because traders would promptly buy, resell, and exploit the opportunities to realize positive returns (Ghose and Yao 2011; Overby and Forman 2015). This process adds liquidity to the market and ensures that prices of the same asset converge. It is how quickly a market responds to flipping opportunities that distinguishes an efficient market from an inefficient one (Malkiel 2003). As such, given a certain number of flipping opportunities, a market where traders can exploit/redress more existing price discrepancies is regarded as more efficient.

Information Transparency Effect

The notion of information transparency refers to how much information is available about an asset at a specific point in time and is usually quantified as the amount of information available to decision-makers (Granados et al. 2010). In accordance with this, we operationalize the level of information transparency of a digital asset in one week as its cumulative amount of blockchain transaction information before that week. According to this operationalization, the length of the trading history governs the level of transparency that traders could have into an asset. For example, longer trading histories can provide traders with greater insight into an asset's bid- ask imbalance, underlying value, and mispricing. We thus establish the impact of the amount of transaction information on market efficiency by comparing the cases with shorter trading histories against those with longer trading histories.

Lack of information transparency hampers market efficiency (Öörni 2003). Numerous studies have documented a positive impact of disclosing richer information on market efficiency. For example, Parker et al. (2016) find that providing timely price information through mobile technologies can increase market efficiency by reducing price dispersion of the crops geographically. In a similar vein, Overby and Forman (2015) find that, by increasing price visibility and reducing transaction costs, electronic commerce helps buyers shift their demands from high-priced places to low-priced ones, thus narrowing price gaps across locations and reducing the likelihood of market failure.

However, the fundamental question of how blockchain's full information transparency impacts market efficiency remains so far unanswered. We attempt to fill this literature gap by re-examining this relationship in the context of blockchain-based marketplaces, where the complete historical trading information is accessible to every trader symmetrically, accurately, instantly, and freely (while this transparency condition is rarely met in other settings). Drawing on the literature, we argue that, at any time point, assets with more historical transaction information are likely to have higher market efficiency compared to those with less transaction information. This is because having additional transaction information can better inform sellers of buyers' potential valuations of the asset, which can rationalize sellers' expectations on the transaction prices, thereby leading to more efficient and uniform asset pricing without huge price discrepancies. Further, from the buyers' perspective, having additional transaction information can reduce their information asymmetry and uncertainty by helping them better forecast future trading prices, thus facilitating efficient asset searching and trading. As such, we posit that at a certain time point, assets with a higher amount of historical transaction information could result in increased market efficiency, as proposed by Hypothesis H1:

H1: At any time point, assets with a higher amount of historical transaction information on the blockchain would have higher market efficiency than those with a lower amount.

Analytical Ability Effect

The importance of users' ability to utilize available information has been well recognized. For example, Rossi and Chintagunta (2016) show that, despite policies aimed at increasing price transparency, over 90% of customers lacked the ability to consume the information of posted fuel price effectively, resulting in persistent price uncertainty and price dispersion. Similarly, Yang et al. (2015) find that full transparency of quote information in the prediction market does not necessarily improve trader performance, and they conjecture that this can be attributed to individuals' cognitive limitations in analyzing and processing the transparent market information.

Our study also closely relates to the burgeoning body of literature on the digital divide. The prior literature on the digital divide primarily focuses on people's unequal access to digital information and technologies across different geographical regions and demographic populations (Burtch and Chan 2019; Dewan and Riggins 2005; Kvasny and Keil 2006). However, a small portion of extant literature has extended the concept of the digital divide to multiple levels, particularly the second level, which pertains to people's heterogeneous ability/skills (i.e., ability divide) to utilize available information resources on the Internet (Wei et al. 2011). Specifically, Wei et al. (2011) propose a three-level digital-divide conceptualization, with the first-level divide being the inequality of access to digital technologies, the second-level being the inequality of the capability to exploit digital technologies and digital information (which is our focus), and the third-level being the inequality of outcomes, such as productivity.

The aforementioned studies, however, primarily investigate the effect of the digital divide on individual-level outcomes, such as individual productivity (Wei et al. 2011) and users' campaign success (Burtch and Chan 2019). Unsurprisingly, few prior works focus on the analytical-ability divide, and most importantly, none of the prior work examines how the deficiency in analytical ability would impact a critical *market-*

level outcome, namely market efficiency. We thus complement the prior literature on ability divide by investigating the significance of "analytical ability" in affecting market efficiency.

Also, notably, most of these extant studies only discuss user analytical ability or ability divide at a conceptual level without an attempt to empirically quantify this key notion using observed data (Rossi and Chintagunta 2016; Yang et al. 2015). Even though a few papers attempt to measure it, their proposed measures are not fully accurate or representative of true capabilities. For example, Wei et al. (2011) use a metric of "computer self-efficacy" captured by the self-evaluations of survey subjects, and similarly, Corgnet et al. (2018) conduct a small-scale lab experiment with the theory of mind (ToM) test to infer traders' trading ability, both of which do not necessarily capture a trader's true analytical ability. In contrast, our study draws on the notion of "analytical ability" from Wei et al. (2011) but, for the first time, quantifies it based on observed historical trading data on the blockchain by leveraging a novel counterfactual-based measure. Our operationalization of traders' analytical ability, to be elaborated in Section 4, critically hinges on the comparison with an ML-augmented counterfactual benchmark, which could effectively rule out the impacts of unobservables and has never been explored in the previous literature regarding analytical ability and its divide (Scheerder et al. 2017; Wei et al. 2011).

According to EMH, an efficient market can only exist if "all market participants agree on the implications of current information for the current price and distributions of future prices of each asset" (Fama 1970). However, due to the heterogeneity of individuals in interpreting historical market information, stemming from their heterogeneous analytical abilities, market efficiency may not be attainable. In other words, even though blockchain-based marketplaces provide all traders with symmetric and complete market information, not everyone benefits equally. High-ability traders, who can better utilize the information, can better forecast the underlying price trend, conduct price discovery, anticipate supply-demand imbalances, and thus exploit more flipping opportunities once they appear. Conversely, traders with low analytical ability may not possess the necessary skills to analyze available data at their disposal, presumably missing profitable flipping opportunities and thereby inhibiting market efficiency. Hence, in a market with high-ability traders, flipping opportunities are more likely to be quickly identified and exploited away. Accordingly, we postulate H2:

H2: A market with higher trader analytical ability has a higher level of market efficiency.

The Moderating Effect of Analytical Ability

While having more information can generally be beneficial, the impact of information transparency on market efficiency may largely depend on traders' analytical ability to utilize the information. This implies that traders' analytical ability may moderate the effect of information transparency on market efficiency, as the potential of historical transaction information can be fully unlocked only when traders possess the necessary skills to effectively consume and utilize the information. For instance, this moderating effect has been observed in the context of IT investment in firms. Jia et al. (2020) show that firms with higher IT capabilities for information automation, standardization, and integration could gain a higher level of efficiency from the implementation of enterprise systems. Likewise, Peukert and Reimers (2022) document a similar moderating effect in the publishing industry. They demonstrate that major publishers who invest more in data-analytics jobs, compared to those investing in text-editing types of jobs, enjoy a higher market efficiency given the same set of publicly available information, such as consumer reviews of books. This finding suggests the crucial role of data analytical ability in moderating the effect of available digital information on market efficiency.

In our case, similarly, higher analytical capabilities could lead to better utilization of market information and, thus, a stronger informational effect in blockchain-based marketplaces. Specifically, given the same amount of information, high-ability traders can quickly identify and exploit existing flipping opportunities by effectively discerning nuanced market conditions and tactfully leveraging the available information to spot short-lived flipping opportunities. On the other hand, low-ability traders may not benefit as much from the available information due to their limited cognitive capacity. Thus, the effectiveness of utilizing blockchain transaction information for decision-making would be restricted for low-ability traders. Taken together, we argue that the blockchain information amount effect would be more pronounced for high-ability traders. We thus posit H3 as follows: **H3**: As traders' analytical ability increases, the effect of blockchain historical transaction information amount on market efficiency becomes stronger.

Data and Variables

Research Context

We choose EnjinX, one of the earliest and leading Ethereum-based NFT marketplaces, as the empirical context. It is among the top ten Metaverse platforms in 2023, according to its market capitalization, with a $100-million investment in the Metaverse virtual space.[2] Backed up by the Ethereum blockchain, EnjinX is a gaming marketplace with its native cryptocurrency Enjin Coin (ENJ) and decentralized ecosystem, where any participant can create, display, sell, and purchase blockchain-based in-game digital items, such as digital weapons and collectibles.

Most importantly, all detailed market trading histories are symmetrically/equally available to everyone through the EnjinX website. EnjinX records and chronologically displays the complete market trading histories for each asset, including trading price, timestamp, as well as change of trader ownerships on its blockchain. The use of blockchain technology ensures that no one can tamper with the asset- or transaction-level information recorded on the blockchain, which provides accurate, complete, and accessible information to all users. As a result, EnjinX traders with the necessary analytical ability should be able to leverage and analyze this transparent data to make informed trading decisions.

In theory, blockchain's symmetric and transparent information structure should lead traders to make similar predictions about future price changes on the same asset. However, in practice, traders may face insufficient transaction information or have imperfect analytical abilities, producing errors in their estimates of price movements. These errors in judgment can cause market inefficiencies to persist, as not all traders can exploit existing price discrepancies (e.g., through flipping) proficiently. If there are flipping opportunities that remain unexploited, the market is inefficient.

We have chosen the JumpNet of EnjinX, a side chain of the Ethereum network, as our research setting. The JumpNet is particularly suitable for our study as it allows for zero Ethereum transaction fees, enabling denser NFT trading frequency compared to the expensive Ethereum Mainnet. This feature creates a more liquid market with richer historical transaction information for traders to conduct price discovery and make informed trades based on their analytical abilities. Further, a necessary condition to achieve market efficiency is minimizing transaction costs (Overby and Forman 2015). The frictionless trading network of JumpNet satisfies this condition better than the Mainnet.

The EnjinX blockchain entails two kinds of information: historical transaction information and asset-specific information. We first use the EnjinX API to collect the transaction data, which consists of the complete trading history of 6,872 gaming assets, including 578,055 listing records and 66,870 sale records made by 123,133 sellers and 1,732 buyers from April 1, 2021, to August 21, 2022, across 74 weeks.[3] Each transaction record contains information on the seller's and buyer's Ethereum addresses, transaction quantity, unit price, timestamp, asset creator, asset name, and the unique asset ID. We also collect public information about each specific asset, including asset fungibility, the number of holders and supplies, transaction fees, and asset properties, which are time-invariant features. Please note that the EnjinX blockchain does not record the game-playing data stored off-chain. In the following subsections, we describe the constructions of our main variables.

Dependent Variable (DV): Market Efficiency

We operationalize market efficiency by following the percentage of exploited flipping opportunities in Subramanian and Overby (2017). It is worth noting that this measurement is applicable to our context by virtue of blockchain's full information transparency on both the *trader-level* and *asset-level* transaction histories. This enables us to identify exactly *who* (by unique trader address) flipped a certain *asset unit* (by

---

[2] See https://coinmarketcap.com/view/metaverse/ and https://www.fxempire.com/forecasts/article/enjin-sets-up-a-100-million-metaverse-fund-the-investments-in-metaverse-space-is-just-starting-798247 (accessed April 05, 2023).

[3] The EnjinX JumpNet was launched in April 2021. It adopts a fixed-price (a.k.a., buy-now price) sale mechanism. Buyers purchase listed assets at the predetermined prices that the sellers post.

unique assetID) during a certain timeframe. In contrast, such trader account-level data is rarely available in traditional stock markets or e-commerce markets.

We first identify the *existence* of flipping opportunities. Specifically, for each sale transaction of an asset, we consider its subsequent sale transactions, occurring in the next α = 24 hours (1 day)[4] for the same asset i in week t, as valid flipping opportunities, if the sale price of the subsequent transaction is higher than that of the earlier transaction plus additional transfer fees. The generic mathematical representation for flipping opportunities is as follows: Transaction k in week t is deemed as a valid flipping opportunity for its preceding Transaction j within the previous α hours, if [5]

$$0.975 \times Price_k - Price_j - TransferFee_k > 0 \qquad (1)$$

The above inequality implies that the buyer of Transaction j could have resold/flipped the same asset for a higher price by making Transaction k, indicating the presence of a flipping opportunity. It is possible that a user purchased an asset intending to use it in game-playing rather than flip it for trading. Such purchases should not count toward valid flipping opportunities, and thus, we only include those asset purchases that appear to have been made with the intent of flipping. Toward that end, we require that the buyer of the preceding transaction j should further post the asset for resale within the next 24 hours to ensure that the buyer does have the intent to flip/resell and that her main objective is to earn returns by trading. As a result, 9,561 sale transactions were followed by a relisting by the same buyers within the next 24 hours. We count future 24-hour flipping opportunities k, only for those 9,561 sales j. Note that we only count each Transaction k uniquely for once to avoid overcounting flipping opportunities. Also, when there are no subsequent transactions within the next 24 hours, or all subsequent transactions' prices are lower than Transaction j's price, there is zero number of flipping opportunities.

As an example, suppose there was a sale associated with a digital sword for $50. We identify three subsequent sales ($80, $60, $40) of the same digital sword sold within 24 hours after the $50 sale. Given the price differences and transfer fees, we conclude that the $80 and $60 sales represented valid flipping opportunities for the $50 sale (but not the $40 sale). That is, there exist two flipping opportunities for the buyer of the $50 sale: she could have flipped/resold the digital sword for higher prices (at $80 and $60). We then aggregate all such flipping opportunities for each asset i in each week t, as the total number of flipping opportunities, $NFlippOpp_{it}$.

Next, we identify the *exploitation* of flipping opportunities, i.e., how many existing inefficiencies have actually been exploited. To do this, out of all identified flipping opportunities, we spot all pairs of transactions of the same asset made by the *same address* (account-level information is now leveraged), i.e., the seller of the subsequent transaction has the same Ethereum address as the buyer of the preceding transaction.[6] Then, we take the latter reselling transaction in each pair of transactions as the exploited ones: by quickly buying, reselling, and making a profit on an asset, a trader address snatches, seizes, and exploits the existing flipping opportunities. Due to this process, the market quickly responds to the presence of flipping opportunities and washes away potential inefficiencies, thereby restoring market efficiency (Subramanian and Overby 2017). We further aggregate all such flipping transactions of asset i in week t as the total number of exploited flipping, $NExploitedOpp_{it}$. To measure the level of market efficiency, we calculate the percentage of exploited flipping opportunities as our dependent variable, which is computed by $PercentExploited_{it} = \frac{NExploitedOpp_{it}}{NFlippOpp_{it}} \times 100$.

As shown in Table 2, the average value of $PercentExploited_{it}$ is 27.37%, suggesting that most existing price discrepancies and market inefficiencies remain unexploited. Hence, the EnjinX marketplace is far from

---

[4] Note that a large portion, about 40% of the resales, happened in 24 hours in our data. As done in Subramanian and Overby (2017), we later vary α for robustness, which yields consistent results.

[5] The transfer fee is designated by NFT creators during the NFT creation process and is paid by sellers to creators per NFT or per transfer in each sale. In addition, EnjinX charges a 2.5% commission fee to sellers from each sale transaction. We thus subtract 2.5% from the price of Transaction k in the first term of Equation 1.

[6] One limitation of our analysis is that we calculate the exploited flips based on individual Ethereum addresses rather than considering each trader's identity. While it is possible that a trader may hold multiple addresses, the blockchain's depersonalized and anonymized nature prevents us from directly observing traders' identities within the data.

being fully efficient. To explain this seemingly paradoxical phenomenon that an information-transparent marketplace unexpectedly experiences substantial inefficiencies, this paper identifies two notable sources that could cause the inefficiencies. We next explain how we operationalize the two inefficiency sources.

Independent Variable: Transaction Information Amount on blockchain

As mentioned earlier, the first inefficiency source stems from the inadequacy in the richness of historical trading information on the blockchain. We capture it by counting *the total number of historical sale transactions* up to week *t* since the creation of asset *i*. This is the independent variable constructed to test H1. Compared to posted prices, realized transaction/sale prices could provide more reliable information on the true valuations of assets (Ghose and Yao 2011), representing more informative and accurate information for traders to utilize. We thus take asset sale history as the main variable to capture the effect of blockchain information transparency and take the total number of listing histories as a control variable.

We next articulate the construction of the other independent variable, trader analytical ability, in a separate section as follows.

Construction of Trader Analytical Ability

Measuring Analytical Ability: Challenges and Our Approach

We first explain why trading performance metrics alone (e.g., annualized return or Sharpe ratio) can *not* be used as the simple proxy for traders' analytical ability. Analytical ability refers to a trader's capability to *analyze* all relevant blockchain information in our context. However, without exploring any available data on the blockchain, a trader may still achieve a satisfactory trading performance by luck or strategies like copy trading, which does not necessarily reflect analytical ability. In other words, trading performance only accounts for trading outcomes, while disregarding the *"analytics"* of the available *information inputs*.

Therefore, to incorporate the analytical aspect into the ability measurement, we propose setting up a cutting-edge ML benchmark. The ML algorithm analyzes the same blockchain information available to human traders, making decisions by utilizing and analyzing transparent market information instead of by luck. This way, in the decisions made by ML, effective analytics of the available blockchain information indeed play a role.

However, it is also problematic to measure analytical ability by directly comparing traders' performance against that of an independent ML algorithm. There are two reasons behind this. First, human traders may possess private knowledge (e.g., game popularity, trading luck) about an asset that is not recorded on the blockchain and cannot be leveraged by ML algorithms. Therefore, in the presence of unobservables that human traders can observe while ML cannot, the predictions of ML may fall behind humans' judgments (Lakkaraju et al. 2017). Accordingly, due to unobservables, directly comparing an independent ML algorithm with human decisions may not necessarily guarantee capturing the true analytical ability of traders.

Second, evaluating the performance of standalone ML models can be infeasible if the ground- truth outcome labels are missing for some of ML's decisions, commonly known as the *"selective labeling problem,"* where the instances that have outcome labels are selectively determined by decision-makers (Kleinberg et al. 2018; Lakkaraju et al. 2017). Specifically, in our context, traders' decision outcomes are selectively recorded and labeled: we can only observe future price changes (e.g., the outcome label) for assets that traders did purchase in reality, but not for assets that traders did not purchase. It is as if the EnjinX blockchain selectively records and presents the price-change outcomes. Hence, if ML makes a "non-recorded" decision that does not have a corresponding ground-truth label (i.e., buying an asset that traders did not buy), assessing the predictive performance of the ML model would be impossible.

To address the above challenges, we argue that a counterfactual scenario is needed, where any unobservables such as luck or copy-trading factors should be ruled out from the measurement, and only humans' analytical capability of leveraging historical market information would be truly counted for the measurement of analytical ability. We, therefore, establish such a counterfactual benchmark by asking, "What if humans had the same analytical capabilities as ML algorithms? What would have happened if humans improved their existing trading strategies by analyzing the blockchain public information as effectively as an ML?" To answer that, we let ML augment traders' existing decisions and examine whether human traders' performance could have been augmented by an ML model with outstanding analytical performance.

The "contraction" method represents one efficient way to realize the ML augmentation (Fu et al. 2021; Kleinberg et al. 2018; Lakkaraju et al. 2017; Shen et al. 2022). We hence adapt and apply the contraction approach to measure the collective analytical ability ($Ability_{it}$) for the group of traders who have purchased asset *i* during week *t*. The principal idea of the contraction method is using ML to augment traders' existing decisions by excluding weak decisions identified by ML. Specifically, the ML algorithm first makes predictions on the same trading decisions that human traders have made, based on the same blockchain data that human traders also have access to. Subsequently, it wittingly sweeps out the suboptimal trading decisions from the original human decision sets based on its predictions, and by doing so, the method effectively improves the quality of traders' existing decisions. This is why the method is called "contraction." And the degree to which the ML contraction can improve (or hinder) traders' performance is used as the barometer to gauge traders' analytical ability.

The contraction method is an ideal solution for us because it effectively resolves the aforementioned challenges. First, the contraction method is the first method that is designed to directly handle the selective labeling problem. It adeptly solves the "non-recorded" decisions by focusing on the "recorded" decision side: suggesting which "recorded" transactions are likely to yield negative returns, and human traders should have avoided them in the original decision sets.

Second, the contraction method effectively addresses the potential impact of unobserved confounders by constructing a quasi-random experimental design. The design assigns human traders with similar trading transactions to the same group, called "cell" (Kleinberg et al. 2018). Within the same group, some traders are assigned to receive the assistance of ML analytics, which will serve as the ML-augmented counterfactual benchmark (i.e., ML + human, the treated group), while the remaining traders facing similar trading decisions trade independently (i.e., control group). By comparing the ML counterfactual benchmark against pure human decisions, our measurement could effectively cancel out the impacts of unobserved factors such as luck or insider information because if those unobservables are really at play, their effects should be parallel and offset in both groups by virtue of the quasi-random design (Kleinberg et al. 2018); the only material difference between the two groups is that the ML algorithm *analytically* augments the treated trader group but not the control group. This way, the contraction method minimizes the potential effects of unobserved confounders in the formation of analytical ability. We also validate the method by ensuring that the quasi-random design is indeed satisfied through the random-assignment test (Kleinberg et al. 2018).

Notably, our paper represents the first attempt to measure analytical ability in a rigorous way where unobserved factors are teased out, and only the analytical aspect is retained in the measurement, through such an ML-augmentation counterfactual benchmark. Without setting up this counterfactual for human ability augmentation, other measurements of analytical ability can be biased or inaccurate (Corgnet et al. 2018). Our study thus makes a methodological contribution by quantifying analytical ability through an innovative approach, in which the counterfactual design of the contraction method is leveraged.

Procedures of the Contraction Approach

For easy reference, we summarize the key steps of the contraction method and the purpose of each step in Table 1, while cutting down the details of the method due to the page limitation.

Estimated Analytical Ability

The contraction method eventually gives rise to the result in Figure 1. The final step quantifies the analytical ability of trader crowds. This is done by examining whether the human quintiles (excluding the 5th quintile point) can outperform the ML-augmented curve in return at the same level of risk. A negative return gap between the human quintiles and the ML-augmented curve suggests that traders still have the potential to enhance their analytical ability by harnessing blockchain's transparent information as efficiently as ML algorithms do. That is, if traders exhibit notable improvements in their return performance through the aid of ML analytics, relative to the cases where they trade independently without such assistance, it signals a substantial opportunity for these traders to amplify their analytical capabilities through the incorporation of ML analytics. Given this idea, if a trader crowd who transacted asset *i* in week *t* is able to outperform the ML-augmented decisions given the same level of risks, $Ability_{it}$ = 1. Otherwise, $Ability_{it}$ = 0.

As shown in Figure 1, human decisions only slightly edge random decisions, whereas the addition of ML assistance can help generate substantially higher returns for traders in quintiles 2, 3, and 4 at the same risk level. Note again that Figure 1 represents an aggregation of all cells, and overall, ML augmentation exhibits better performance than human traders alone. However, it is possible for traders to occasionally outperform ML augmentation in some individual cells, which is why we also observe trader crowds with high analytical ability. As a result, 30% (70%) of our observations are of high (low) analytical ability.

Steps	Tasks	Purposes	Descriptions
1	Define the problem	Confirm the problem objective for traders.	Generating positive returns by buying assetsis the main objective of traders. Calculate the actual realized return for each purchase decision.
2	ML training	Establish the ML analytics-based performance, which is generated completely based on information analytics.	Train XGBoost as the benchmark with the complete historical input data; ML then predicts the likelihood of receiving a positive return by making a purchase decision.
3	Form cells	To create a quasi-random experimental design for the ML counterfactual, group similar trading decisions in one cell as if they are randomly assigned.	Assign trading instances into different cells based on the combination of the top four most important asset-level features.
4	Form trader crowds	Eliminate biases, such as trading luck, by grouping shared decision-making crowds who traded the same asset in a week.	Within each cell, combine transactions of each asset in each week as the decisions of a trader crowd.
5	Form trader quintiles	Account for heterogeneous risk preferences of traders for the risk-return trade-off.	Within each cell, divide trader crowds into five equal quintiles. Then, calculate the total return and total risk of human transactions in each quintile, which represents the actual human performance (Kleinberg et al. 2018).
6	ML contraction	Let ML augment human's existing decisions and thus construct the ML-augmented counterfactual for the next-step comparison with pure human traders.	ML augments the 5th quintile trader crowds' decisions by sequentially removing weak decisions from the 5th quintile, based on its predicted return probability, from the lowest to highest probability sequence.
7	Estimate trader analytical ability	See what would have happened if humans in quintiles 1, 2, 3, and 4 had followed ML's analytics-based trading strategies.	Compare the total return of the ML-augmented counterfactual curve with that of human decisions in quintiles 1, 2, 3, and 4, under the same risk level.

Table 1. Overview of the Contraction Approach

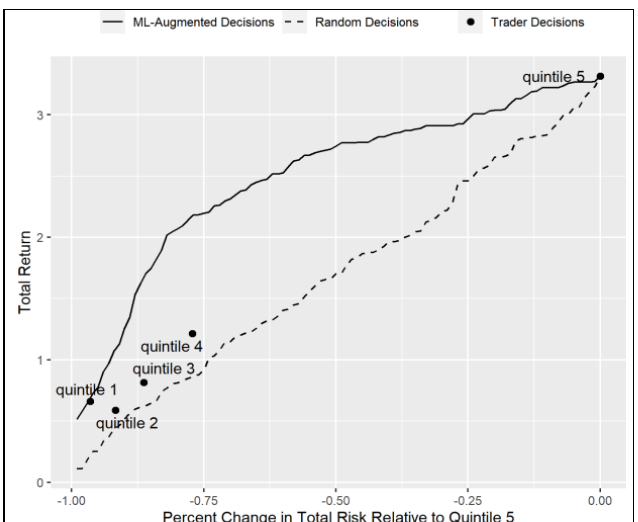

Figure 1. Contraction Result

Empirical Strategy and Results

Empirical Model and Main Results

Next, we test the main effects of transaction information amount (H1) and users' analytical ability (H2) on market efficiency by adopting the following two-way fixed effects specification:

$$Y_{it} = \beta_1 AssetSaleHistory_{it} + \beta_2 Ability_{it} + \beta_3 Controls_{it} + \upsilon_i + u_t + \epsilon_{it} \quad (2)$$

We estimate the above Specification 2 by ordinary least squares (OLS) (where $Y_{it} = PercentExploited_{it}$) and fractional logit and probit models (where $Y_{it} = PercentExploited_{it}/100$, as the fractional logit/probit models require the DV to be between 0 and 1). β_1 is of main interest for testing H1, capturing the informational effect of blockchain sale history on market efficiency. β_2 is of main interest for testing H2, capturing whether and how much the high analytical ability of the trader crowd who transacted asset i in week t could effectively improve market efficiency. We further include several asset-level control variables to mitigate the problem of omitted variables. As mentioned earlier, we first control for the total number of historical listings of asset i before week t. Second, we control for how long asset i has been created (i.e., asset tenure) up to week t to represent the maturity of asset i, to rule out the possibility that β_1 reflects the effect of asset maturity. Third, we use the average historical sale price of asset i before week t to control for the potential influence of asset prices on future flips. Finally, because flipping opportunities and realized flips in the current week t essentially stem from the listing activities in week t (i.e., people have to list the asset before making a flip), we also include the total number of listings of asset i in week t to control for asset trading opportunities. We present the definition and descriptive statistics of the key variables in Table 2.

Variables	Descriptions	Mean	Std. Dev.
Dependent Variable			
$PercentExploited_{it}$	The percentage of exploited flipping opportunities of asset i in week t	27.37	32.38
$NExploitedOpp_{it}$	The total number of exploited flipping opportunities of asset i in week t	1.27	3.08
$NFlippOpp_{it}$	The total number of flipping opportunities of asset i in week t	6.84	16.50
Independent Variable			
$AssetSaleHistory_{it}$	The total number of sale transactions of asset i before week t	162.0	327.38
$Ability_{it}$	The analytical ability of the trader crowd who transacted asset i in week t; Equal to 1 if the trader group can outperform the cutting-edge ML	0.31	0.46
$AssetSaleHistory \times Ability$	The interaction between information amount and analytical ability	40.05	132.20
Control Variable			
$AssetListingHistory_{it}$	The total number of historical listings of asset i before week t	2068.11	3858.68
$AssetTenure_{it}$	The number of days since asset i's creation till week t	26.90	61.78
$AvgSalePrice_{it}$	The average historical sale price of asset i before week t	3.90	21.77
$Nlisting_{it}$	The number of listings of asset i in the current week t	237.55	495.89

Table 2. Summary Statistics

We also add the asset fixed effects (υ_i) into the specification to capture the unobserved asset- level time-invariant characteristics (e.g., asset reserved inherent value defined by its creator, asset functionalities in game playing). u_t controls for the week fixed effects, allowing us to better isolate common time-specific shocks, such as ENJ price fluctuations and the overall market dynamics. ϵ_{it} denotes the error term. Throughout the paper, we cluster the standard errors at the asset level.

The results for H1 and H2 are reported in Table 3 Columns (1), (3), and (5). Both coefficients of asset sale history and ability are significant and positive, lending support to H1 and H2. More specifically, in OLS, given everything else being equal, having ten additional historical transaction information can significantly increase market efficiency by 1.10% (=0.03/27.37*10*100%, where 27.37 is the average market efficiency level), and trader groups with high analytical ability can exploit 69.02% (=18.89/27.37*100%) more flipping opportunities compared to low-ability groups.

Further, to test H3, we examine the moderation effect between the two main variables with the following specification:

$$Y_{it} = \beta_1 AssetSaleHistory_{it} + \beta_2 Ability_{it} + \beta_3 AssetSaleHistory_{it} \times Ability_{it} + \beta_4 Controls_{it} + v_i + u_t + \epsilon_{it} \quad (3)$$

β_3 is of main interest here, capturing to which degree the effect of blockchain information on market efficiency can be intensified (or dampened) if the information is utilized by high-ability traders. The result for the moderation effect is shown in Table 3 Columns (2), (4), and (6). We observe a significantly stronger impact of blockchain historical information (specifically an increase of 0.11% in the informational effect) when traders are more capable of handling transparent market information. The two main effects (H1 and H2) are consistent. We then use $Y_{it} = PercentExploited_{it}$ for all following analyses.

Variable	OLS		Fractional Logit		Fractional Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
AssetSaleHistory	0.03^{**} (0.014)	0.03^{**} (0.014)	0.003^{**} (0.001)	0.003^{**} (0.001)	0.001^{**} (0.0007)	0.002^{**} (0.001)
Ability	18.89^{***} (3.156)	13.47^{***} (3.716)	1.30^{***} (0.213)	0.93^{***} (0.243)	0.74^{***} (0.116)	0.53^{***} (0.135)
AssetSaleHistory × *Ability*	-	0.03^{***} (0.009)	-	0.002^{***} (0.0005)	-	0.001^{***} (0.0003)
Controls	YES	YES	YES	YES	YES	YES
Asset Fixed Effects	YES	YES	YES	YES	YES	YES
Week Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	507	507	507	507	507	507
Adjusted R^2	0.36	0.38	-	-	-	-
Notes. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors are clustered at the asset level.						

Table 3. Main Results

Identification

Since our study uses observational data without experimental random variation, endogeneity may be present. We next resolve the potential endogeneity concerns resulting from three specific problems: omitted variables, selection bias, and measurement error. For each problem, we offer our empirical strategy to strengthen the identification and show that our main results are robust and unlikely to be biased.

Omitted Variables: FEct and IVs

The existence of some omitted variables for our two key independent variables in the econometric specifications may confound the identification of the main analysis. The first approach we embrace to alleviate the omitted-variable issue is "fixed effects counterfactual estimator" (FEct) (Liu et al. 2022). We apply FEct by adding two-way fixed effects with jackknife estimates over 500 iterations. One notable limitation of FEct is that it cannot accommodate interaction terms currently. Thus, in an attempt to capture the moderating effect, we set up two subsamples where the values of *AssetSaleHistory* differ. Specifically, we select two cutoffs to construct the two subsamples: observations with at least 1 historical sale and 21 historical sales (which constitute 86% and 70% of the full sample, respectively). We then run FEct separately based on the three samples, including the full sample and the two subsamples. We expect that, if the moderating effect really holds, the effect of analytical ability will be stronger when the number of historical sales goes higher. As shown in Table 4, our main findings of H1 and H2 still hold for all these three samples. Moreover, the coefficient magnitudes of $Ability_{it}$ increase from Column (1) to (3), as the sample contains observations with more blockchain sale histories. This finding, therefore, aligns with the moderating effect of H3, lending support to our result robustness under the FEct estimation.

Variable	Full Sample	Sample with Sale History≥1	Sample with Sale History≥21
	(1)	(2)	(3)
AssetSaleHistory	0.05*** (0.018)	0.04*** (0.013)	0.05** (0.024)
Ability	11.00** (5.402)	38.87*** (4.669)	45.41*** (5.077)
Controls	YES	YES	YES
Asset Fixed Effects	YES	YES	YES
Week Fixed Effects	YES	YES	YES
Optimal Method	IFEct, $r^* = 1$	IFEct, $r^* = 1$	MC

Notes. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Jackknife standard errors are in parentheses. r represents the optimal number of unobserved factors (Liu et al. 2022).

Table 4. Address the Endogeneity of Omitted Variables: FEct

We further employ the instrument variable (IV) estimation to account for endogeneity of *AssetSaleHistory*. Particularly, there can exist unobserved information and factors (e.g., individual idiosyncratic trading preferences and asset popularity) that are correlated with *AssetSaleHistory* and may also influence market efficiency but are nevertheless omitted in the specifications. In general, a valid IV should satisfy two presumptions: (1) IV relevance, meaning that IVs should be highly correlated with the endogenous variables, and (2) IV exogeneity, meaning that IVs should be uncorrelated with the error term and should not affect the outcome variable directly.

We first construct one IV for *AssetSaleHistory*, which is the total number of melting histories of asset i before week t, *AssetMeltHistory*. Melting a piece of an asset would permanently destroy the asset and make it unavailable for sale. The rationale of this IV is that an asset with more melting history is likely to have fewer historical sales because more melting history indicates that people tend to place low valuations on the asset, manifested by melting and devaluing the asset. It is also possible that *AssetMeltHistory* can be positively associated with *AssetSaleHistory*, as higher *AssetMeltHistory* may indicate that the asset possesses excessive copies (oversupply), which further leads to more trading possibilities and thus higher number of sale transactions. It is, however, very unlikely that asset melting history would directly affect market efficiency since market efficiency is mainly driven by market sales/flips instead of those melting records. In other words, it can only affect market efficiency indirectly through assets' historical sales.

We next construct two IVs for $Ability_{it}$. The first one is $MintPerSaleHistory = \frac{\text{Number of Assets Minted by the Creator}}{\text{Number of Assets Sold by the Creator} + 1}$, defined as the average number of assets minted/created by the *creator* of asset i in order to get one sold before week t, excluding the focal asset i. It essentially proxies how competent the focal asset's creator was before week t, at minting assets and successfully selling them. We next elaborate on why this IV is valid. First, this IV is strongly and negatively related to $Ability_{it}$. If the focal asset is minted by an incompetent creator who needs to produce lots of different assets before eventually making a sale (i.e., *MintPerSaleHistory* is high), this asset would unlikely be transacted by high-ability traders in the first place (i.e., $Ability_{it}$ would be low), because the asset and its creator are inherently unattractive (i.e., low-quality creators are unlikely able to allure high-ability traders). Meanwhile, this IV can also be strongly and negatively correlated with *AssetSaleHistory* because a low-quality creator who cannot sell many of her creations may have fewer transaction histories for the focal asset as well. However, this IV does not directly influence market efficiency because market efficiency is primarily determined by the reselling/flips by secondary traders of the focal NFT creation, whereas the instrument only accounts for the first selling activity of all other creations made by the original creator, which does not contribute to the flipping and market efficiency of the focal asset.

The second instrument for $Ability_{it}$ is $Ability_{-it}$, which represents the average analytical ability level of other assets (excluding focal asset i) in the same week. The rationale is that the analytical ability of the trader crowds who traded other assets could be highly correlated with the analytical ability of the focal trader crowd who traded the focal asset in the same week (they can potentially be trading partners). However, the analytical abilities of people trading other assets are unlikely to directly impact the focal asset's market efficiency.

This way, we construct three IVs in total for the two endogenous independent variables. We then run 2SLS (two-stage least squares) for H1 and H2. The result of the second-stage estimation is displayed in Table 5 Column (1). Table 5 also presents the Kleibergen-Paap rk Wald F-statistic, which confirms that the IVs are indeed strong, and therefore the IV relevance assumption passes (*F*-statistic > 10% critical value). Further, to statistically check the validity of the IV exogeneity assumption, we perform the Hansen-*J* over-identification test, where the *p*-value suggests that we cannot reject the null hypothesis that the three IVs are correctly excluded from the main specification. Thus, the IV exogeneity assumption also holds.

For the moderating effect, we further use four IVs, which are the interactions between the three original IVs. We thus end up having 7 IVs in total to estimate Specification (3). The 2SLS result of H3 is displayed in Table 5 Column (2). The two tests still pass, and overall, 2SLS gives rise to consistent results.

Variable	(1)	(2)
AssetSaleHistory	0.06^{**} (0.028)	0.09^{***} (0.029)
Ability	18.64^{***} (3.823)	13.41^{***} (4.122)
AssetSaleHistory × *Ability*	-	0.02^{**} (0.010)
Controls	YES	YES
Asset Fixed Effects	YES	YES
Week Fixed Effects	YES	YES
Kleibergen-Paap rk Wald F statistic	20.85	34.09
Stock-Yogo 10% maximal IV size	13.43	8.50
Hansen J statistic *p*-value	0.12	0.95
Observations	504	504
Centered R^2	0.15	0.16
No. of Excluded IVs	3	7
Notes. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors are clustered at the asset level.		

Table 5. 2SLS with IVs

Selection Bias: CEM and Heckman Two-Step Correction

The second endogeneity concern stems from the self-selection of high- vs. low-ability traders. We thus implement the CEM approach to make the assets traded by these two types of traders more similar to each other (Iacus et al. 2012). We first adopt a one-to-one matching method, where each observation transacted by a high-ability trader crowd (i.e., a treated unit) is matched with another similar observation transacted by a low-ability trader crowd (i.e., a control unit). Based on the matched sample and the CEM weights, we re-estimate Specifications 2 and 3 for the three hypotheses. The results presented in Table 6 Columns (1) and (2) are consistent with the main results. We also try one-to-multiple matching based on the same set of matching characteristics, and the three hypotheses and our main findings still hold.

Variable	**CEM One-to-One Matching**		**SIMEX**	
	(1)	(2)	(3)	(4)
AssetSaleHistory	0.04^{***} (0.015)	0.04^{**} (0.015)	0.02^{**} (0.007)	0.01^{*} (0.000)
Ability	20.26^{***} (3.728)	15.14^{***} (4.762)	12.49^{***} (1.569)	10.59^{***} (1.732)
AssetSaleHistory × *Ability*	-	0.03^{***} (0.009)	-	0.03^{***} (0.009)
Controls	YES	YES	YES	YES
Asset Fixed Effects	YES	YES	YES	YES
Week Fixed Effects	YES	YES	YES	YES
Observations	408	408	507	507
Adjusted R^2	0.32	0.33	-	-
Notes. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors are clustered at the asset level.				

Table 6. Address the Endogeneity of Self-Selection and Measurement Error: CEM, SIMEX

Our estimation may suffer from another selection bias due to the fact that the dependent variable, *PercentExploited*, is observed only for observations where the number of flipping opportunities is positive, but not for those with zero flipping opportunities. It may happen that these two groups of observations are systematically different, rendering the outcome variable not missing at random. We thus resort to the Heckman two-step selection model, which is commonly used to correct the bias resulting from non-randomly selected samples (Ayabakan et al. 2017; Heckman 1979). The results corrected by the Heckman two-step approach do not significantly change our main findings, further confirming that the selection bias does not represent a severe threat to our identification (results omitted).

Measurement Error: SIMEX

The measurement error problem may also be present since no ML prediction is perfect. The analytical ability variable generated from the contraction method may thus suffer from measurement errors due to misclassification, which can bias the subsequent econometric estimations (Qiao and Huang 2021; Yang et al. 2018). To correct the measurement errors underlying the analytical ability variable, we adopt the simulation extrapolation method (SIMEX), which is documented as an effective method to tackle measurement errors in Yang et al. (2018). The results of SIMEX in Table 6 Columns (3) and (4), again, confirm that the coefficient sizes and directions are highly consistent with the main results.

Robustness Checks

In this subsection, we carry out several robustness checks that pertain to our outcome variable, market efficiency (results omitted). First, in the main analysis, we take $\alpha = 24$ as the time interval for flipping. Here, we also vary the value of α with 12 (half days), 72 (three days), and 120 (five days) hours and recalculate the percentage of exploited flipping opportunities. We then rerun the specifications based on the new market-efficiency values. The results are highly consistent with the main analysis. Another potential concern is that the cryptocurrency price can fluctuate a lot, even within one day. To get a more consistent comparison of prices for flipping and returns, we map each sale price in ENJ to its corresponding USD closing price on the day of sale. We then redo the analysis based on USD (including the measurement of market efficiency and the return metrics of analytical ability), and the results suggest the robustness of our findings. Finally, we ensure the robustness of our results by taking *NExploitedOpp* as the DV while controlling for a given number of existing flipping opportunities. As the DV is a count number, we also estimate Specifications 2 and 3 using the negative binomial and Poisson models. We obtain qualitatively similar results.

Discussion and Conclusion

The blockchain technology is designed to provide complete, accurate, and transparent market information, which should ideally lead to full market efficiency as per Fama (1970) and Granados et al. (2012). However, we uncover and empirically demonstrate a new source of market inefficiency in the context of blockchain-based markets, which stems from traders' inability to analyze blockchain transparent information.

Theoretical Implications

First and foremost, this study contributes to the literature on information economics and market efficiency by challenging the conventional belief that full information transparency can ultimately eliminate market inefficiencies. We discover a fundamental paradox that market inefficiencies persist due to traders' incompetence in their analytical ability to leverage blockchain transparency. Further, our paper contributes to the digital-divide literature by suggesting that the nature of the digital divide transitions to the form of the capability divide when information is ubiquitous and accessible to all. Now that blockchain technology has bridged the digital access chasm, a new type of divide, namely the analytical-ability divide, warrants attention. This paper answers the call of Wei et al. (2011) to investigate the second-level digital divide in the form of the analytical-ability divide.

We also make a methodological contribution by adapting the contraction method to quantify traders' analytical ability in a two-sided market. We present a novel use case of the contraction method in measuring analytical ability by creating a quasi-random design to compare traders' actual performance against the ML-augmented analytical counterfactual. Although scholars have recognized the importance of consumers' ability to analyze the available transparent information, none of them attempted to rigorously quantify it (Corgnet et al. 2018; Rossi and Chintagunta 2016; Yang et al. 2015), partially due to data limitations. For instance, in traditional stock trading or e-commerce settings, neither traders nor consumers have free or easy access to complete account-level transaction history, let alone quantify analytical ability and examine its impact on market efficiency. As such, our work fills the literature gap by measuring analytical ability using blockchain's full transparency of history and effectively accounting for the existence of unobservables and the selective labeling issue through the contraction approach, which represents an efficient lever to reveal traders' analytical abilities.

Managerial Implications

Achieving high market efficiency is crucial for platforms because a higher percentage of flipping exploitation could generate higher commission revenue. Hence, from a managerial perspective, our study suggests that blockchain-based platforms should consider effective designs to bridge the analytical-ability divide, in order to subsequently help traders to identify and exploit any available flipping opportunities, and thus boost platform revenues. Toward that aim, blockchain-based marketplaces should provide not only transparent and accessible historical market information but also necessary analytical tools to assist traders' decision-making, e.g., ML-generated analyses for traders to better utilize the transparent information and gain deeper insights into the data. This implication, highlighting the necessity of effective analytics of available market information, echoes the findings in (Peukert and Reimers 2022), where digitization and analytics improve market efficiency by aiding publishers in predicting book sales and facilitating resource reallocation. It is not a coincidence that some NFT platforms have already taken steps to incorporate analytical tools for traders to use. Platforms such as OpenSea.co and Rarible.com have integrated preliminary solutions such as price distribution graphs and NFT scarcity indicators. In fact, there are even platforms specifically designed for NFT sale analytics, such as the Dune analytics dashboard and DappRadar.com ranking. These real-world practices lend support to our proposition of narrowing the "analytical-ability divide." While digital information technology is readily available to all, the challenge lies in making users more proficient in their utilization of the available information. In the meantime, individual traders should also realize the importance of effectively utilizing transparent data to their advantage, preferably with the assistance of ML. Finally, our findings also shed light on the implications for regulators. To ameliorate market efficiency and increase overall social welfare, regulators may re-think how to regulate and incentivize two-sided platform owners to provide effective analytical tools to market participants.

Concluding Remarks

Overall, our study provides important insights into the challenges faced by blockchain-based markets. By shifting the focus from the traditional digital access divide to the new analytical-ability divide in an information-symmetric setting, our study highlights the importance of individual capabilities in affecting market efficiency. We argue that the analytical-ability divide is currently superseding the digital access divide as the new significant challenge to emerging blockchain-empowered markets. We conclude with the following quote.

> *"The digital revolution can open doors to unseen opportunities and industries, but only if everyone has access to the internet and the ability to use it."*
>
> \- Melinda Gates, co-founder of the Bill & Melinda Gates Foundation.

References

Akerlof, G. A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics* (84:3), pp. 488-500.

Ayabakan, S., Bardhan, I. R., and Zheng, Z. 2017. "A Data Envelopment Analysis Approach to Estimate It-Enabled Production Capability," *Management Information Systems Quarterly* (41:1), pp. 189-206.

Bayer, P., Geissler, C., Mangum, K., and Roberts, J. W. 2020. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market," *The Review of Financial Studies* (33:11), pp. 5212-5247.

Burtch, G., and Chan, J. 2019. "Investigating the Relationship between Medical Crowdfunding and Personal Bankruptcy in the United States: Evidence of a Digital Divide," *Management Information Systems Quarterly* (43:1), pp. 237-262.

Corgnet, B., Desantis, M., and Porter, D. 2018. "What Makes a Good Trader? On the Role of Intuition and Reflection on Trader Performance," *The Journal of Finance* (73:3), pp. 1113-1137.

Dewan, S., and Riggins, F. J. 2005. "The Digital Divide: Current and Future Research Directions," *Journal of the Association for Information Systems* (6:12), pp. 298-337.

Fama, E. F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work," *The Journal of Finance* (25:2), pp. 383-417.

Fu, R., Huang, Y., and Singh, P. V. 2021. "Crowds, Lending, Machine, and Bias," *Information Systems Research* (32:1), pp. 72-92.

Ghose, A., and Yao, Y. 2011. "Using Transaction Prices to Re-Examine Price Dispersion in Electronic Markets," *Information Systems Research* (22:2), pp. 269-288.
Granados, N., Gupta, A., and Kauffman, R. J. 2010. "Research Commentary—Information Transparency in Business-to-Consumer Markets: Concepts, Framework, and Research Agenda," *Information Systems Research* (21:2), pp. 207-226.
Granados, N., Gupta, A., and Kauffman, R. J. 2012. "Online and Offline Demand and Price Elasticities: Evidence from the Air Travel Industry," *Information Systems Research* (23:1), pp. 164-181.
Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica: Journal of the Econometric Society*), pp. 153-161.
Huang, N., Burtch, G., He, Y., and Hong, Y. 2022. "Managing Congestion in a Matching Market Via Demand Information Disclosure," *Information Systems Research* (33:4), pp. 1196-1220.
Iacus, S. M., King, G., and Porro, G. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis* (20:1), pp. 1-24.
Jia, N., Rai, A., and Xu, S. X. 2020. "Reducing Capital Market Anomaly: The Role of Information Technology Using an Information Uncertainty Lens," *Management Science* (66:2), pp. 979-1001.
Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. 2018. "Human Decisions and Machine Predictions," *The Quarterly Journal of Economics* (133:1), pp. 237-293.
Kvasny, L., and Keil, M. 2006. "The Challenges of Redressing the Digital Divide: A Tale of Two Us Cities," *Information Systems Journal* (16:1), pp. 23-53.
Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., and Mullainathan, S. 2017. "The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables," *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 275-284.
Liu, L., Wang, Y., and Xu, Y. 2022. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data," *American Journal of Political Science*).
Malkiel, B. G. 2003. "The Efficient Market Hypothesis and Its Critics," *Journal of Economic Perspectives* (17:1), pp. 59-82.
Öörni, A. 2003. "Consumer Search in Electronic Markets: An Experimental Analysis of Travel Services," *European Journal of Information Systems* (12), pp. 30-40.
Overby, E., and Clarke, J. 2012. "A Transaction-Level Analysis of Spatial Arbitrage: The Role of Habit, Attention, and Electronic Trading," *Management Science* (58:2), pp. 394-412.
Overby, E., and Forman, C. 2015. "The Effect of Electronic Commerce on Geographic Purchasing Patterns and Price Dispersion," *Management Science* (61:2), pp. 431-453.
Parker, C., Ramdas, K., and Savva, N. 2016. "Is It Enough? Evidence from a Natural Experiment in India's Agriculture Markets," *Management Science* (62:9), pp. 2481-2503.
Peukert, C., and Reimers, I. 2022. "Digitization, Prediction, and Market Efficiency: Evidence from Book Publishing Deals," *Management Science* (68:9), pp. 6907-6924.
Qiao, M., and Huang, K.-W. 2021. "Correcting Misclassification Bias in Regression Models with Variables Generated Via Data Mining," *Information Systems Research* (32:2), pp. 462-480.
Rossi, F., and Chintagunta, P. K. 2016. "Price Transparency and Retail Prices: Evidence from Fuel Price Signs in the Italian Highway System," *Journal of Marketing Research* (53:3), pp. 407-423.
Scheerder, A., Van Deursen, A., and Van Dijk, J. 2017. "Determinants of Internet Skills, Uses and Outcomes. A Systematic Review of the Second-and Third-Level Digital Divide," *Telematics and Informatics* (34:8), pp. 1607-1624.
Shen, Z., Zheng, E., and Jiang, W. 2022. "A Machine Learning Approach to Mitigating Irrationality in Copy Trading," *Available at SSRN 4298755*).
Subramanian, H., and Overby, E. 2017. "Electronic Commerce, Spatial Arbitrage, and Market Efficiency," *Information Systems Research* (28:1), pp. 97-116.
Wei, K.-K., Teo, H.-H., Chan, H. C., and Tan, B. C. 2011. "Conceptualizing and Testing a Social Cognitive Model of the Digital Divide," *Information Systems Research* (22:1), pp. 170-187.
Yang, M., Adomavicius, G., Burtch, G., and Ren, Y. 2018. "Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated Via Data Mining," *Information Systems Research* (29:1), pp. 4-24.
Yang, S., Li, T., and van Heck, E. 2015. "Information Transparency in Prediction Markets," *Decision Support Systems* (78), pp. 67-79.