

Are Digital Infrastructures Ready to Scale?

Lead-Lag Relationships in Market Microstructure

Relationship between Customers' Size of Wallet, Share of Wallet, and Total Wallet: New Insights from the Blockchain

A Synergistic Alliance – Will Generative AI Help Tackle the ESG Data Problem?



## **Impressum**

Redaktion  
Prof. Dr. Peter Gomber  
Julian Schmidt, M.Sc.

Herausgeber  
Prof. Dr. Peter Gomber  
Stellvertretender Vorstandsvorsitzender  
efl – the Data Science Institute e.V.

Kontakt  
[info@eflab.de](mailto:info@eflab.de)  
[www.eflab.de](http://www.eflab.de)

Gestaltung  
Novensis Communication GmbH  
Weilburg

1. Ausgabe, 2024  
Auflage: 200 Stück (Print) / 2.000 Stück (E-Paper)

Copyright © by efl – the Data Science Institute e.V.

Printed in Germany  
ISSN 1866-1238

## Editorial

# Are Digital Infrastructures Ready to Scale?

Stephanie Eckermann

Yes, they are. The securities services industry is at a tipping point of its digital transformation and will now see industry solutions to scale. We identify three fundamental drivers being adopted more broadly: cloud migration, data, and digitization. This triage also drives the scaling of Clearstream's digital infrastructure D7, one of the leading digital infrastructures globally.

At Clearstream, technology at scale is at the core of our business. We hold around 40% of all Eurobonds as well as around 10% of all fixed-income assets globally in our custody, and we settle 1.2 million trades each day. More than half of all T2S transactions go through our books.

This kind of scale has been delivered for decades to our clients and our supply chain partners by our own servers on premise ensuring the resilience and high levels of efficiency that are critical to us. Still, we see this

technological context to change now and digital infrastructures to scale. And the triage of cloud migration, data, and digitization will drive this development.

As for many other companies, cloud is the basis of our digital transformation at Clearstream. Though the cloud journey has its challenges, from a technical and a regulatory point of view, our strong conviction remains that, without cloud, unlocking the benefits of data and AI as well as digitization would prove much more challenging. Apart from the rich tooling coming with it, cloud enables us to improve our time-to-market capabilities significantly.

In 2018, we started to move our on-premise hosting to the cloud on a phased basis, starting with development and functional testing. In 2021, we moved to acceptance testing and this year to non-core production. Our initial goals, increased agility and resilience, have been achieved.



**Dr. Stephanie Eckermann**  
Chief Executive Officer  
Clearstream Banking AG

The second driver besides cloud clearly is data. Having migrated much of our infrastructure to the cloud, we wanted to unlock the enormous data benefits as an early deliverable. We started the development of our Clearstream data layer in 2021, delivering data analytics for our business and for our clients.

We transmit around 2 billion pieces of information per year to our clients – an average of almost 8 million per day. It is an enormous data and data transport challenge, and the volume keeps increasing. Our data lake currently holds 40 TB of data volume, yet it is still in its infancy. We are increasing capacity at 0.5 TB per day. To support this, we have established over 160 ETL (extract, transform, and load) pipelines to load data from on-premise to the cloud.

The biggest challenge of our industry though is not the ability to process large amounts of data, but to work with high-quality data. And here the third driver, digitization, plays an

important role to set digital standards with clean sets of data.

While digital assets might be considered as being in a more muted market phase due to the "crypto winter", we are seeing increased interest in successful digitization use cases that offer both the benefits of digital transactions and a regulated environment. With D7 as our digital infrastructure, we invest in use cases to digitize assets in regulated markets together with our clients – be it retail structured products under German eWpG or bonds under MiCA in the EU. Thus, we are one of the leading digital infrastructures in securities services globally with over 7,000 digital issuances and a volume of more than EUR 3 billion to date.

Finally, our conviction is that the triage of cloud, data, and digitization will herald a new era of financial services innovation, both in digital securities and in data services.

## Research Report

# Lead-Lag Relationships in Market Microstructure

THIS STUDY EXPLORES HIGH-FREQUENCY CROSS-ASSET LEAD-LAG RELATIONSHIPS FOR VARIOUS MARKET MICROSTRUCTURE DIMENSIONS. UTILIZING DATA FROM STOCKS, FUTURES, AND EXCHANGE TRADED PRODUCTS, THE FINDINGS UNCOVER SIGNIFICANT LEAD-LAG PATTERNS, PARTICULARLY AMONG FUNDAMENTALLY RELATED INSTRUMENTS. OUR RESULTS DEMONSTRATE THAT KNOWLEDGE ABOUT LEAD-LAG RELATIONSHIPS CAN BE LEVERAGED FOR FORECASTING SHORT-TERM CHANGES IN FINANCIAL MARKETS.

Micha Bender

Julian Schmidt

### Introduction

In financial markets, lead-lag effects refer to situations where some financial instruments provide indications of future price, liquidity, or volatility developments in other instruments. These effects arise due to unequal information dissemination across markets.

The emergence of algorithmic and high-frequency trading (HFT) has led to a significant shift in information dynamics within financial markets. In this fast-paced environment, order book activities, such as order submissions or cancellations, rather than trades have become the central source of information (Brogaard et al., 2019). High-frequency traders (HFTs) utilize

Tino Cestonaro

real-time market data and can react within less than a millisecond to changes in the order book. Their actions are often driven by patterns within the order book dynamics, rather than relying on fundamental information. These pattern-based trading activities of HFTs have led to increased co-movements in returns and liquidity, driven by the rapid dissemination of information and correlated HFT strategies (Melcaniec et al., 2019).

This transformation in the trading behavior raises the question of whether lead-lag relationships are more related to market microstructure information, such as order book activities, rather than to changes in fundamental information (Huth and Abergel, 2014). Understanding

these changing relationships entails both opportunities and risks for various market participants. Changes in lead-lag relations may expose institutional investors and market makers to the risk of unknown interdependencies within their portfolios. However, these changes can also present opportunities for HFTs.

### Data and Methodology

To investigate the existence of pattern-driven lead-lag relationships in financial markets, we analyze non-contemporaneous correlations across a broad spectrum of assets and market microstructure dimensions like prices, liquidity, or volatility. Thereby, we follow a data-driven approach by examining all potential lead-lag relationships and do not rely on any underlying assumptions such as the fundamental relationship between two instruments, e.g., derivatives and their underlying. To conduct the analysis, we use trade and order book data with nanosecond precision from Deutsche Börse's A7 Analytics Platform, encompassing 19 heavily traded

financial instruments, such as the EURO STOXX 50 Future (FESX), the Bund-Future, or the Xetra-Gold exchange-traded commodity. Our sample period spans six months, from January 2021 to June 2021. To quantify and assess lead-lag relationships, we utilize the estimator introduced by Hayashi and Yoshida (2005). This so-called HY-estimator estimates the correlation between two irregularly spaced time series (X and Y) with different lengths. The HY-estimator is calculated for a discrete grid of leads and lags. In our analysis, this grid covers leads and lags from a few microseconds up to one minute. The resulting HY-curve then consists of the estimated correlation coefficients ( $\hat{\rho}$ ) for each predefined lead or lag  $l$ .

In Figure 1, we show the HY-curve for representative instruments pairs from different asset classes. For example, the black-dotted HY-curve shows the estimated correlation coefficients for the midpoints of the DAX Future (FDAX) and the transaction prices of the DAX exchange-traded

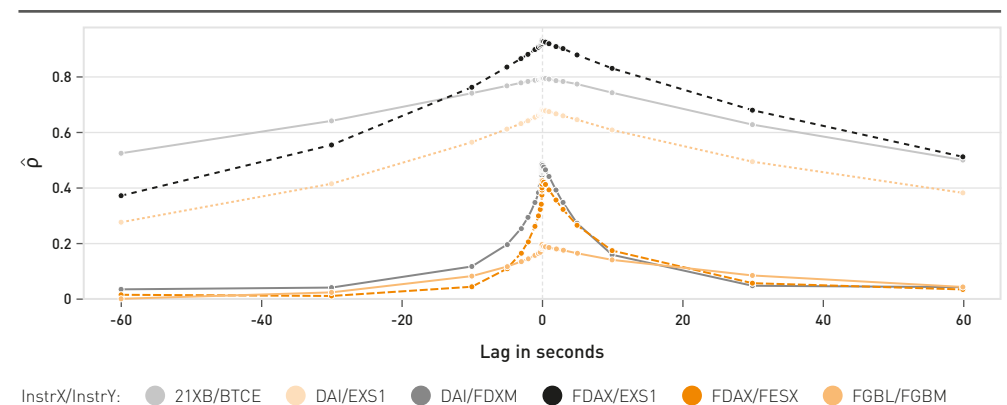


Figure 1: HY-Curves for X: Midpoint and Y: Transaction Price for Selected Instruments

fund (EXS1) with a maximum correlation of more than 0.8 for lags between +/- 10 seconds.

**Results**

Our results uncover significant cross-asset lead-lag relationships among various microstructure measures. In order to assess the magnitude of the detected lead-lag relationships, we introduce a novel measure called lead-lag-strength (LLS) that takes the skewness of the HY-curve and the overall correlation into account. Measuring lead-lag relationships with the LLS, we find the strongest relationships between transaction prices and midpoints, the order book imbalance (OBI) and midpoints, and between transaction prices and midpoint volatility. In particular, our results imply that changes in the OBI and the transaction price in a given asset provide information about future changes in the midpoint of another asset, which can also be seen from Figure 1 where almost all HY-curves are skewed to the right. These lead-lag patterns vary among instruments, but predominantly

exist among financial instruments that are fundamentally related, such as those with common underlying components (e.g., the FDAX and FESX) or those exposed to the same risk factors (e.g., long-term and medium-term bond futures).

Instrument-wise, we find the strongest lead-lag relationship between FESX's transaction prices leading FDAX's midpoints (see also the dark orange-dotted curve in Figure 1). On average, the reaction of the FDAX takes less than a microsecond, highlighting the speed of information transmission in modern financial markets. In addition, we provide evidence that the detected lead-lag effects are persistent over time and do not depend on the current level of liquidity or volatility. Regarding the determinants of a lead-lag relationship, our results show that instruments with higher liquidity and trading activity tend to lead other instruments in almost all microstructure measures examined.

Lastly, our results reveal the predictive power of

the identified lead-lag relationships using the LLS. We evaluate this predictive power by utilizing the leading (in terms of the LLS) time series to predict the direction of the next change in the lagging time series. The prediction accuracy for each instrument combination within the selected measure combinations of this LLS-based prediction approach is displayed as a dot in Figure 2. Furthermore, the LLS value itself is shown for each instrument combination along with descriptive statistics for the LLS-based prediction approach and a naïve benchmark. Overall, the LLS emerges as a valuable indicator for identifying lead-lag relationships and can, e.g., be used to develop statistical arbitrage trading strategies in HFT.

**Conclusion**

The speed of information dissemination and the increasing pattern-driven trading behavior raise the question whether lead-lag relationships are more related to market microstructure information, such as limit order book activities, rather than to changes in the (fundamental) price of an asset. To tackle this question, this study examines cross-asset lead-lag relationships between different microstructure dimensions using data from Xetra and Eurex. We find high-frequency lead-lag effects for price, OBI, and volatility measures among fundamentally related instruments. Our findings indicate that short-term lead-lag effects in the HFT environment still embed fundamental information,

albeit at an accelerated pace and derived from the most granular pieces of information, such as a single order submission. These results help market operators and regulators to anticipate and assess potential contagion effects and threats to market efficiency and stability. Additionally, our findings can enhance the development of protective mechanisms that consider various microstructure measures and operate across different instruments. Professional investors can leverage the identified lead-lag relationships using the LLS to predict changes in lagging assets.

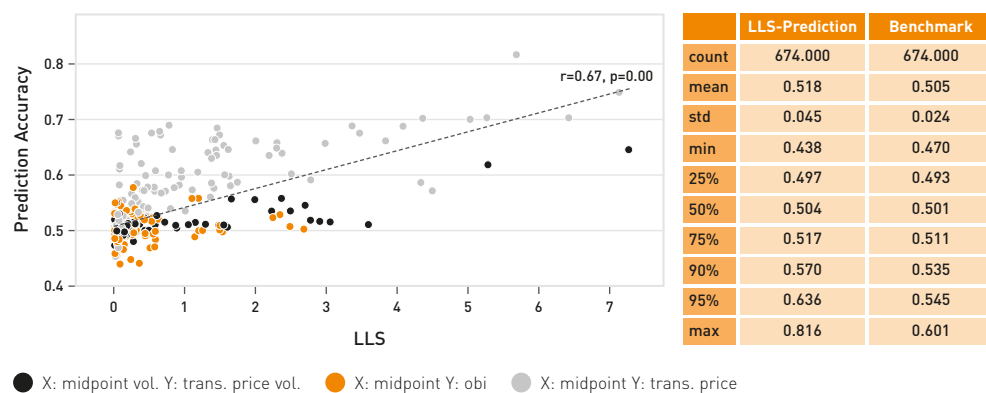
**References**

**Brogaard, J.; Hendershott, T.; Riordan, R.:** Price Discovery without Trading: Evidence from Limit Orders. In: Journal of Finance, 74 (2019) 4, pp. 1621–1658.

**Huth, N.; Abergel, F.:** High Frequency Lead-Lag Relationships – Empirical facts. In: Journal of Empirical Finance, 26 (2014), pp. 41–58.

**Hayashi, T.; Yoshida, N.:** On Covariance Estimation of Non-synchronously Observed Diffusion Processes. In: Bernoulli, 11 (2005) 2, pp. 359–379.

**Malceniene, L.; Malceniaks, K.; Putninš, T. J.:** High Frequency Trading and Comovement in Financial Markets. In: Journal of Financial Economics, 134 (2019) 2, pp. 381–399.



**Figure 2: Relationship between Average Prediction Accuracy (based on LLS) and the LLS**

## Research Report

# Relationship between Customers' Size of Wallet, Share of Wallet, and Total Wallet: New Insights from the Blockchain

NOWADAYS, FIRMS LACK INFORMATION TO DERIVE THE SHARE OF WALLET, A VITAL METRIC THAT IDENTIFIES HOW MUCH ADDITIONAL SPENDING A FIRM COULD CAPTURE FROM EACH CUSTOMER. HOWEVER, DECODING BLOCKCHAIN DATA ENABLES OBSERVING ALL TRANSACTIONS OF EACH WALLET, RESPECTIVELY CUSTOMER, ON THE ETHEREUM NFT MARKET. TO SHED LIGHT ON THE SHARE OF WALLET, WE ANALYZED 22.7 MILLION TRANSACTIONS FROM OVER 1.3 MILLION CUSTOMERS ACROSS EIGHT COMPETING FIRMS ON THE ETHEREUM NFT MARKET.

Björn Hanneke

Bernd Skiera

Thilo Gerwien Kraft

Oliver Hinz

### Introduction

Customers' share of wallet is the share of a customer's total spending that goes to a particular firm within a specific market. It is a valuable metric because it identifies how much additional spending a firm could capture from each customer. Information about the size of wallet, i.e., the revenue or profit contribution of a customer, is readily available in firms that observe their customers' behavior over time, such as online retailers, banks, energy suppliers, and telecommunication providers. Unfortunately, it is much more difficult to derive the share of wallet because a customer's total wallet is typically not

observable in traditional settings (Chen and Steckel, 2012). This unavailability of important data limits the ability to derive conclusions about the relationships between share of wallet, size of wallet, and total wallet.

Remarkably, the emergence of Blockchain technology and its continued adoption introduces new possibilities to explore and empirically investigate the relationships between the share and the size of wallet. Blockchains store transaction information publicly on distributed ledgers (Peres et al., 2023), and decoding them enables us to analyze transactions, firms, and

customers on the Ethereum non-fungible token (NFT) market with unprecedented accuracy and low costs. This study is the first that exploits the unique opportunities that decoding of the Blockchain provides to derive valuable metrics in customer relationship management (CRM), here mainly the share of wallet.

Our empirical analysis focuses on the share and size of wallet. However, the fundamental argument of this article is that the opportunities that come with decoding Blockchains go far beyond an analysis of the share and size of wallet. Therefore, we also describe how Blockchain transaction transparency affects other parts of CRM, particularly those regarding customer acquisition and retention.

One of the important findings of the study is the significant variation among companies in terms of the distribution of their customers across quintiles and their effectiveness in securing a large share of their customers' wallet, indicating their competitive positioning within the NFT market.

### Research Setting

An NFT is a cryptographic asset with a unique identifier, differentiating it from cryptographic fungible tokens, e.g., cryptocurrencies (Peres et al., 2023).

Our study focuses on NFTs as tokenized digital assets, where ownership transfers are achievable by transferring the NFT. Hence, the NFTs of

our study represent heterogeneous digital goods whose creators can sell them in primary markets, and subsequent owners can re-sell them in secondary markets. Examples include digital collectibles, such as PORSCHE 911 NFTs, digital sneakers for metaverses, e.g., by Nike, and virtual land in the metaverse, e.g., The Sandbox.

Like traditional firms, Blockchain-based firms earn fees from successful transactions, e.g., by charging sellers a percentage of the transaction volume. The main difference between Blockchain-based and traditional firms is that traditional firms store transactions on internal databases. In contrast, Blockchain-based firms rely on Blockchains to store transactions. To interact with the Blockchain, Blockchain-based firms deploy smart contracts on the Blockchain (Peres et al., 2023). Customers interact with the firm's smart contract to settle transactions, e.g., process payments, transfer assets, and store transaction outcomes on the Blockchain (Cong and He, 2019).

The interaction with smart contracts requires a cryptographic wallet, e.g., to initiate or sign transactions. These cryptographic wallets, subsequently referred to as wallets, are identifiable by their unique wallet address (Jørgensen and Beck, 2022).

In 2022, when cryptocurrency prices rose to new heights, NFT trading peaked in terms of trading volumes in USD. Since its peak in 2022, falling

cryptocurrency prices and decreasing transaction numbers led to falling NFT trading volumes. Even though NFT trading is an emerging industry, Blockchain-based firms for NFT trading have generated significant fees and facilitated millions of transactions between NFT owners (Borri et al., 2022).

### Description of the Empirical Study

For our empirical study, we leverage Ethereum data on NFT sales to investigate customers' size and share of wallets across eight firms. We focus on data from Ethereum because it pioneered NFT trading and is more mature than other Blockchains.

We retrieve our data from Flipside Crypto, a major Blockchain data provider. Flipside Crypto offers various Blockchain data by curated data models, thus offering stable results and public, accessible documentation. The documentation

allows tracking and verifying data decoding activities back to the raw transaction records. Furthermore, Flipside Crypto's NFT sales data allows us to select customers that sold NFTs at one of eight major Ethereum-based firms, i.e., Blur, GEM, LooksRare, NFTX, OpenSea, Rarible, sudoswap, and X2Y2. For every customer and firm, we retrieve the number of transactions and fees in USD, hence the size of wallets. As firms charge fees to sellers, we focus our analysis on sales exclusively. Our sample includes all NFT sales between January 1, 2022, and December 31, 2022.

Across the eight firms (see Table 1), we observe 1.86 million customers, which conduct 22.79 million transactions, paying over USD 1.02 billion in fees, representing most of the firm-enabled Ethereum NFT trading market during our observation horizon, e.g., excluding peer-to-peer transactions.

Firms	Number of Customers	Fees (USD m)	Fees per Customer (USD)	Market Share (Fees)	Number of Transactions (USD k)	Transactions per Customer	Market Share (Transactions)
OpenSea	1,211,026	454.0	374.9	44.3%	19,650.7	16.2	86.2%
GEM	305,727	5.6	18.5	0.6%	1,545.3	5.1	6.8%
Blur	112,596	0.5	4.8	0.1%	523.0	4.6	2.3%
X2Y2	103,409	17.3	166.9	1.7%	602.5	5.8	2.6%
LooksRare	87,053	546.1	6,273.3	53.2%	277.5	3.2	1.2%
sudoswap	24,805	0.4	14.1	0.0%	125.6	5.1	0.6%
Rarible	1,629	1.2	116.9	0.1%	40.3	3.8	0.2%
NFTX	5,404	0.0	3.0	0.0%	23.8	4.4	0.1%
<b>Sum, Mean</b>	<b>1,860,649</b>	<b>1,025.1</b>	<b>871.5</b>	<b>100.0%</b>	<b>22,788.7</b>	<b>6.0</b>	<b>100.0%</b>
<b>HHI</b>				<b>4,802</b>			<b>7,496</b>

Note: m for million, k for thousand, HHI for Herfindahl-Hirschman Index

Table 1: Summary Statistics for All Firms on the Ethereum NFT Market

### Results of the Empirical Study

Summary statistics on firms and customers describe a market that is, to a degree, concentrated on a few firms and customers (see Table 1).

Considering the number of firms alone, the NFT market's HHI values (Herfindahl-Hirschman Index for the estimation of market concentration) of 4,802 for market share (fees) and 7,496 for market share (transactions) reflects that the market is still young and maturing with some firms capturing a large portion of market shares.

Table 2 summarizes the firms' successes in capturing a large share of wallets across their customers. OpenSea's customers have a significant share of wallet (>80%) with this firm, an exception in the market. However, some smaller firms have noticeable amounts of cus-

tomers with a large share of wallets, e.g., sudoswap (64.8%) and Rarible (52.1%), and to a lesser extent LooksRare (19.9%). These findings might indicate that these firms' offerings differ from those of other firms.

The distribution of shares of wallets across firms reflects that most customers sell on a few platforms, e.g., OpenSea and one of the other firms. This behavior results in a high portion of customers with a high share of wallet at OpenSea and a high portion of customers with a low share of wallet with the other firms. OpenSea's prominent market position is also apparent by exploring customers' transactions across firms. Regarding fees, OpenSea is only second.

For comparison, we have included the mean values for the share of wallet estimates from a major U.S. bank investigated by Du et al. (2007).

Firms	Share of Wallets					Sum	Mean
	0-20%	20-40%	40-60%	60-80%	80-100%		
OpenSea	0.6%	0.8%	1.6%	3.3%	93.6%	100.0%	96.1%
GEM	79.0%	7.7%	4.0%	1.8%	7.5%	100.0%	15.5%
Blur	90.7%	3.7%	1.9%	0.8%	2.9%	100.0%	7.1%
X2Y2	79.3%	5.9%	3.3%	1.7%	9.8%	100.0%	16.7%
LooksRare	61.3%	10.0%	5.3%	3.4%	19.9%	100.0%	30.4%
sudoswap	27.3%	2.9%	2.7%	2.2%	64.8%	100.0%	69.1%
Rarible	35.1%	5.9%	4.0%	3.0%	52.1%	100.0%	59.0%
NFTX	95.4%	0.6%	0.5%	0.4%	3.1%	100.0%	4.0%
<b>Mean</b>	<b>58.6%</b>	<b>4.7%</b>	<b>2.9%</b>	<b>2.1%</b>	<b>31.7%</b>	<b>100.0%</b>	
<b>Du et al. (2007)</b>	<b>54.7%</b>	<b>13.9%</b>	<b>9.3%</b>	<b>7.6%</b>	<b>14.6%</b>	<b>100.0%</b>	

Table 2: Firms' Customer Distribution over Share of Wallet Baskets for Ethereum NFT Market

As their study estimated total wallets to derive the share of wallets, we expect differences. However, there are similarities, such as the overall distribution, e.g., the 20% basket being the largest, followed by the largest share of wallet basket (>80%). Also, there is a decreasing trend in the share of customers from the 20% basket to the 80% basket. However, this large difference is primarily due to OpenSea's strong market position.

In summary, there are notable differences between firms regarding distributions of their customers (quintiles) and their success in capturing a large share of wallet of their customers, reflecting their competitive positioning in the NFT market. We suggest distinguishing observed firms into market leaders, followers, and niche players.

The growth potential for the market leader OpenSea is somewhat limited, as most customers already have a significant share of wallets with OpenSea. Nonetheless, the outstanding position of OpenSea presents an opportunity for other firms to engage in competitive actions and gain customers from OpenSea. However, market followers and niche players seem to take different competitive approaches.

The niche players, Rarible and sudoswap, seem to have successfully attracted different customer groups, as both firms have customers with large shares of wallets. On average, Rarible generated over USD 116 in fees per customer,

suggesting a comparably attractive but small niche (1,629 customers and USD 1.2 million fees). In contrast, sudoswap has more customers (24,805) but is less profitable overall (USD 0.4 million) and per customer (USD 14.1).

The market followers GEM, Blur, X2Y2, LooksRare, and NFTX share similar distributions across their customer bases and share of wallets. Notably, they all seem to focus on large customers, which are overrepresented across their customer bases. Therefore, they seem to follow the traditional marketing advice to focus on customers with large total wallets.

Furthermore, we investigate correlations between firms' observable variables, such as the size of wallet and the number of transactions, and unobservable variables, such as total wallet and share of wallet. We hardly observe correlations between firms' observable and unobservable variables.

#### Further Research in Customer Relationship Management

Firms may use the total wallet and potential wallet metrics to operationalize their CRM goals along the CRM activities of customer acquisition, development, and retention.

Fungible and non-fungible tokens play a key role as marketing instruments for Blockchain-based firms. Firms may emit tokens to finance their operations (Cong et al., 2019), use tokens as means of payment within their firm, platform, or

ecosystem (Nofer et al., 2017), and relate token holdings to decentralized governance mechanisms (Beck et al., 2018).

Tokens have exciting applications for CRM activities for three reasons: First, they constitute a tradeable value. Second, they are transferrable with little friction and costs. Third, they are available to Blockchain-based firms at marginally no cost.

However, data privacy is a major issue, as over time wallets will potentially be identifiable, either by enough data within a wallet or by linking off-chain data to a wallet's transactions. As such, it is questionable if observable solutions adhere to existing privacy regulations, such as the GDPR (Skiera et al., 2022).

#### References

- Beck, R.; Müller-Bloch, C.; King, J. L.:** Governance in the Blockchain Economy: A Framework and Research Agenda. In: *Journal of the Association for Information Systems*, 19 (2018) 10, pp. 1020–1034.
- Borri, N.; Liu, Y.; Tsyvinski, A.:** The Economics of Non-Fungible Tokens. Working Paper, 2022.
- Chen, Y.; Steckel, J. H.:** Modeling Credit Card Share of Wallet: Solving the Incomplete Information Problem. In: *Journal of Marketing Research*, 49 (2012) 5, pp. 655–669.

**Cong, L. W.; He, Z.:** Blockchain Disruption and Smart Contracts. In: *Review of Financial Studies*, 32 (2019) 5, pp. 1754–1797.

**Du, R. Y.; Kamakura, W. A.; Mela, C. F.:** Size and Share of Customer Wallet. In: *Journal of Marketing*, 71 (2007) 2, pp. 94–113.

**Jørgensen, K. P.; Beck, R.:** Universal Wallets. In: *Business & Information Systems Engineering*, 64 (2022) 1, pp. 115–125.

**Nofer, M.; Gomber, P.; Hinz, O.; Schiereck, D.:** Blockchain. In: *Business & Information Systems Engineering*, 59 (2017) 3, pp. 183–187.

**Peres, R.; Schreier, M.; Schweidel, D. A.; Sorescu, A.:** Blockchain Meets Marketing: Opportunities, Threats, and Avenues for Future Research. In: *International Journal of Research in Marketing*, 40 (2023) 1, pp. 1–11.

**Skiera, B.; Miller, K.; Jin, Y.; Kraft, L.; Laub, R.; Schmitt, J.:** The Impact of the General Data Protection Regulation (GDPR) on the Online Advertising Market. Self-Publishing, Frankfurt am Main, 2022.



## Insideview

# A Synergistic Alliance – Will Generative AI Help Tackle the ESG Data Problem?

INTERVIEW WITH TOMAS VAN DER HEIJDEN

**In a year marked by challenging market dynamics, the importance of ESG investments remains unwavering. But the wave of ESG regulations and requests generates a demand for more scalable ways to collect and analyze ESG data. The rise of AI could mark a turning point in an industry heavily burdened by reporting requirements, and unlock the true potential of ESG for businesses and investors alike.**

**What is the significance of ESG data in today's business landscape?**

ESG data allows businesses to assess risks related to environmental impact, social responsibility, and governance practices. ESG metrics are integral in investment decisions. At the same time, regulatory bodies worldwide are implementing strict ESG regulations, making it mandatory for businesses to disclose and document their ESG performance. That's why robust and granular data is crucial.

Because ESG is so broad, collecting ESG data requires multiple departments and organisations to collaborate more closely than ever before.

**What are some key challenges businesses and financial market participants face when it comes to working with ESG data?**

ESG data can be extremely diverse, and it often lives in unstructured data sources (such as PDFs, excel sheets, email chains). In addition, ESG reporting is still far from a standardised practice; many regulatory bodies and rating agencies rely on their own methodology, making ESG metrics challenging to compare and analyze. Finally, the sheer volume of ESG data to collect can be overwhelming.

**How are businesses tackling this endeavour? Why is Generative AI a potential solution?**

Many businesses are still figuring out things



Tomas van der Heijden  
CEO and Co-Founder  
Briink

as they go. Broadly speaking, large corporate players are building their internal capacities, often spearheaded by in-house teams, but they also rely on third-party data providers. Medium and small businesses are still lagging behind. This is why Generative AI has such a huge potential in this space. There is a global shortage of human capital (e.g. ESG analysts) to enable companies to extract and analyze this data on their own, and outsourcing is expensive. AI can process large volumes of textual or visual ESG data and generate meaningful insights at scale.

**How is this helping ESG knowledge workers today?**

ESG is about so much more than gathering and analyzing data – it's about decision-making. AI empowers professionals to focus on strategy and stakeholder engagement, rather than solely on data collection and analysis.

**What do you see this field evolving into and what interesting trends do you see applying here?**

We have really only scratched the surface of what AI can do. In the future, AI for ESG will be capable of integrating more diverse data sources, like tables and graphs. In addition, we will witness a gradual shift away from "generalist" AI (like chatGPT), and a proliferation of "small language models" that are optimized for specific use-cases, like ESG. At Briink, we have tested the effectiveness of such approaches through domain-specific fine-tuning, prompt engineering and retrieval augmented generation. This is crucial in an industry where there is a growing emphasis on trust and transparency, where concerns about greenwashing have persisted over the past decade.

**Thank you for this interesting conversation.**

# Infopool

## News

### Annual Conference 2023 of the efl was a Great Success

The efl Annual Conference on “Opportunities and Challenges of Generative AI” took place on November 15<sup>th</sup>, 2023. More than 300 participants gained insights from leading researchers, practitioners, and regulators on how AI is revolutionizing different areas of society and economy such as finance, marketing, and education. Speakers also delved into the critical challenges of Generative AI, including privacy concerns and market concentration. Deutsche Bundesbank as a co-host provided its unique perspective on the subject. We thank Prof. Dr. Bernd Skiera and his team for organizing this year’s conference!

### Academic Fellow of Marketing Science Institute

Prof. Dr. Bernd Skiera was appointed as an Academic Fellow of Marketing Science Institute (MSI), which is MSI’s highest honor. Congratulations!

### Federal Ministry of Economics and Climate supports project led by Prof. Dr. Hinz

The project “Consumer Empowerment und Teilhabe in der Domäne Smart Living” (COMET) led by Prof. Dr. Oliver Hinz and his team is supported by the Federal Ministry of Economics and Climate (BMWK). The aim of the project is to enable data donation from smart living devices and apps for consumers who can on their side benefit from sophisticated reports and decision support. The project started in November 2023 and is funded with around EUR 900,000.

### Prof. Dr. Hinz Took over the Position of the Editor-in-Chief of BISE

Prof. Dr. Oliver Hinz took over the position of the Editor-in-Chief of the renowned journal “Business & Information Systems Engineering (BISE)”, formerly known as “Wirtschaftsinformatik”. It is one of the most influential scientific journals in the Information Systems (IS) discipline in Europe and its first issue was published in 1959. Congratulations!

### Niklas Trimpe Joined the efl in November 2023

Niklas Trimpe studied Business Informatics at Goethe University and subsequently joined the Chair of e-Finance (Prof. Dr. Peter Gomber). His research focuses on machine learning applications in financial markets and market microstructure related topics. Welcome to the efl!

### Awards for Excellent Teaching and Dissertation Supervision

Prof. Dr. Andreas Hackethal and Dr. Benjamin Clapham achieved 1<sup>st</sup> places in the teaching evaluation of the Faculty of Economics and Business at Goethe University in the summer term 2023 for their lectures in the Bachelor and Master program, respectively. Prof. Dr. Bernd Skiera received an award from the Goethe Research Academy for Early Career Researchers as the Best Dissertation Supervisor of Goethe University Frankfurt. Congratulations!

### Best Paper Runner-Up at DEEM’23

The paper “DiffML: End-to-end Differentiable ML Pipelines” by Benjamin Hilprecht and co-authors under the supervision of Prof. Dr. Carsten Binnig was awarded as Best Paper Runner-Up at DEEM’23. Congratulations!

## Selected efl Publications

### Ahmadi, I. ; Abou Nabout, N.; Skiera, B.; Maleki, E.; Fladenhofer, J.:

Overwhelming Targeting Options: Selecting Audience Segments for Online Advertising. Forthcoming in: International Journal of Research in Marketing.

### Bauer, K.; Heigl, R.; Hinz, O.; Kosfeld, M.:

Feedback Loops in Machine Learning: A Study on the Interplay of Continuous Updating and Human Discrimination. Forthcoming in: Journal of the Association of Information Systems.

### Carl, K. V.; Mihale-Wilson, C.; Zibuschka, J.; Hinz, O.:

A Consumer Perspective on Corporate Digital Responsibility – An Empirical Evaluation of Consumer Preferences. Forthcoming in: Journal of Business Economics.

### Cestonaro, T.; De Paolis, J.; Panz, S.:

High-Frequency Price Formation in Fragmented Equity Markets. In: 2023 Financial Management Association Annual Meeting (FMA); Chicago (IL), US, and 63<sup>rd</sup> Annual Meeting of the Southern Finance Association (SFA); Fajardo, Puerto Rico, 2023.

### Clapham, B.; Jakobs, J.; Schmidt, J.; Gomber, P.; Muntermann, M.:

A Taxonomy of Violations in Digital Asset Markets. In: Proceedings of the 44<sup>th</sup> International Conference on Information Systems (ICIS); Hyderabad, India, 2023.

### Grahl, J.; Rothlauf, F.; Hinz, O.; Abdel-Karim, B.; Mihale-Wilson, C.:

How do Likes Influence Revenue? A Randomized Controlled Field Experiment. In: Journal of Business Research, 167 (2023), Article No. 114133.

### Hanneke, B.; Baum, L.; Schlereth, C.; Hinz, O.:

Consumer Preferences for Privacy Management Systems. In: Proceedings of the 44<sup>th</sup> International Conference on Information Systems (ICIS); Hyderabad, India, 2023.

### Thostrup, L.; Doci, G.; Boeschen, N.; Luthra, M.; Binnig, C.:

A Distributed GPU Joins on Fast RDMA-Capable Networks. In: Proceedings of the ACM on Management of Data, 1 (2023) 1, Article No. 29.

### Ziegler, T.; Nelson-Slivon, J.; Leis, V.; Binnig, C.:

Design Guidelines for Correct, Efficient, and Scalable Synchronization Using One-Sided RDMA. In: Proceedings of the ACM on Management of Data, 1 (2023) 2, Article No. 131.

For a comprehensive list of all efl news postings and efl publications see: <https://www.eflab.de>

# Infopool

## RESEARCH PAPER:

### NAVIGATING THE JAGGED TECHNOLOGICAL FRONTIER: FIELD EXPERIMENTAL EVIDENCE OF THE EFFECTS OF AI ON KNOWLEDGE WORKER PRODUCTIVITY AND QUALITY

How much more productive does Generative Artificial Intelligence make knowledge workers? Based on this research question, the authors assess the real-world impact of Generative Artificial Intelligence on consultants' productivity and task quality in complex environments. The paper's key message is that Artificial Intelligence can significantly enhance the productivity and quality of work in knowledge-intensive tasks, but its effectiveness varies depending on the nature of the task. For tasks within the Artificial Intelligence's capabilities, workers using Artificial Intelligence (especially GPT-4) perform faster and with higher quality. However, reliance on Artificial Intelligence can hinder performance for tasks beyond the Artificial Intelligence's current capabilities.

**Dell'Acqua, F.; McFowland, E.; Mollick, E.; Lifshitz-Assaf, H.; Kellogg, K; Rajendran, S; Krayer, L.; Candelon, F.; Lakhani, K.**

**In: Harvard Business School Technology & Operations Management Unit Working Paper No. 24-013.**

## RESEARCH PAPER:

### CAN FOUNDATION MODELS WRANGLE YOUR DATA?

Foundation Models (FMs) are models trained on large corpora of data that, at very large scale, can generalize to new tasks without any task-specific finetuning. This paper aims to understand an underexplored area of FMs: classical data tasks like cleaning and integration; and it evaluates the performance of FMs on these tasks. The authors find that large FMs generalize and achieve impressive performance on data cleaning and integration tasks on a few selected datasets, even though the FMs are not trained for these data tasks. However, a more representative study is needed to get the full picture of the FM's capabilities on data tasks. In particular, numeric data sets are under-represented in the study.

**Narayan, A.; Chami, I.; Orr, L.; Ré, C.**

**In: Proceedings of the VLDB Endowment, 16 (2023) 4, pp. 738-746.**

## efl insights

The efl publishes the insights in the form of a periodic newsletter which appears two times a year. Besides a number of printed copies, the efl insights is distributed digitally via E-mail for reasons of saving natural resources. The main purpose of the efl insights is to provide latest efl research results to our audience. Therefore, the main part is the description of two research results on a managerial level – complemented by an editorial, an interview, and some short news.

For receiving our efl insights regularly via E-mail, please subscribe on our homepage [www.eflab.de](http://www.eflab.de) (> NEWS > SIGN UP EFL INSIGHTS) as we need your E-mail address for sending the efl insights to you. Alternatively, you can mail your business card with the note "efl insights" to the subsequent postal address or send us an E-mail.

**Prof. Dr. Peter Gomber**  
**Vice Chairman of the efl – the Data Science Institute**  
**Goethe University Frankfurt**  
**Theodor-W.-Adorno-Platz 4**  
**D-60629 Frankfurt am Main**

[insights@eflab.de](mailto:insights@eflab.de)

**Further information about the efl is available at**  
**[www.eflab.de](http://www.eflab.de).**



The efl – the Data Science Institute is a proud member of the House of Finance of Goethe University, Frankfurt.  
For more information about the House of Finance, please visit [www.hof.uni-frankfurt.de](http://www.hof.uni-frankfurt.de).

THE EFL – THE DATA SCIENCE INSTITUTE IS AN INDUSTRY-ACADEMIC RESEARCH PARTNERSHIP BETWEEN FRANKFURT AND DARMSTADT UNIVERSITIES AND PARTNERS DEUTSCHE BÖRSE GROUP, DZ BANK GROUP, AMERICAN EXPRESS, DEUTSCHE BUNDESBANK, AND FINANZ INFORMATIK LOCATED AT THE HOUSE OF FINANCE, GOETHE UNIVERSITY, FRANKFURT.

For further  
information  
please  
contact:

Prof. Dr. Peter Gomber  
Vice Chairman of the  
efl – the Data Science Institute  
Goethe University Frankfurt  
Theodor-W.-Adorno-Platz 4  
D-60629 Frankfurt am Main

**Phone** +49 (0)69 / 798 - 346 82  
**E-mail** [gomber@wiwi.uni-frankfurt.de](mailto:gomber@wiwi.uni-frankfurt.de)

**Press contact**  
**Phone** +49 (0)69 / 798 - 346 82  
**E-mail** [presse@eflab.de](mailto:presse@eflab.de)

**or visit our website**  
<http://www.eflab.de>