Julia Kreppmeier

Digitization in the Financial Industry: Empirical Evidence on Data Privacy and Digital Assets

University of Regensburg
Regensburg, Germany

Digitization in the Financial Industry: Empirical Evidence on Data Privacy and Digital Assets

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg

vorgelegt von: Julia Kreppmeier

Berichterstatter:
Prof. Dr. Gregor Dorfleitner
Prof. Dr. Klaus Röder

Tag der Disputation:
20.07.2023

Julia Kreppmeier

Digitization in the Financial Industry: Empirical Evidence on Data Privacy and Digital Assets

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

Advisors:
Prof. Dr. Gregor Dorfleitner
Prof. Dr. Klaus Röder

University of Regensburg
Regensburg, Germany

Acknowledgements

First of all, I owe my gratitude to my supervisor, Prof. Dr. Gregor Dorfleitner, for his constant advice, encouragement, and the opportunities he provided me during my time at the Department of Finance at the University of Regensburg. Above all, I am grateful for his trust and for persuading me to do my Ph.D. in the first place. I am also thankful to Prof. Dr. Klaus Röder for being my second advisor. The inspiring discussions with both of them and their valuable suggestions have profoundly improved my research.

Moreover, I would also like to thank my colleagues at the Department of Finance for constructive discussions and suggestions. Furthermore, I owe thanks to the Deutsche Bundesbank for financially supporting the project "Analyse des FinTech-Marktes in Deutschland, Fortführung der bestehenden Aktivitäten, Erweiterung um Webinhalte inkl. quantitativer Analysen" (project no. 7208857).

My special thanks also go to my family and friends for their continuous encouragement, support, and patience.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, the advent of digitization has brought significant changes to all areas of life. Notably, the financial industry has experienced a profound transformation due to advances in information technology. This phenomenon is mainly attributable to financial products being often entirely information-based and devoid of physical components or interactions, such as online payments as opposed to car purchases (Puschmann, 2017).

In the course of this, new market entrants and emerging technologies are challenging the established supremacy of traditional financial institutions as financial intermediaries. Most of these innovations predominantly impact financial intermediation through two distinct channels. Firstly, platforms facilitate matching between projects and investors, standardize information, and provide a streamlined path to complete investments. Secondly, technologies such as Distributed Ledger replace financial intermediaries and enable total disintermediation (Bollaert et al., 2021). In response to the challenges posed by digitization in the financial sector, the European Commission (2020) adopted the Digital Finance Package. It aims to promote a competitive financial industry in the European Union that provides consumers with access to innovative financial services while addressing risks related to consumer protection and financial stability.

The acronym FinTech is often encountered in connection with digitization in the financial industry and is derived from the combination of the two words *financial* and *technology*. There is no single legal definition that encompasses and is unanimously agreed upon by all perspectives of the stakeholders in the financial sector. A critical feature underlying numerous definitions is that FinTech firms leverage technology to augment financial services (Deutsche Bundesbank, 2023; Chen et al., 2019; European Banking Authority, 2017; Financial Stability Board, 2017; Schueffel, 2016). Therefore, it has become customary to delineate FinTechs based on their business activities and, thus, their business model.

Throughout this thesis, the FinTech taxonomy of Dorfleitner et al. (2017) is applied, graphically displayed in Figure 1.1. The overarching segmentation into asset management, financing, payments, and other FinTechs corresponds to the major stages of the value chain of traditional banks and financial service providers. Chen et al. (2019) provide empirical evidence that the Internet of Things, robo advice, and blockchain are the most valuable FinTech innovations for the entire financial sector.[1]

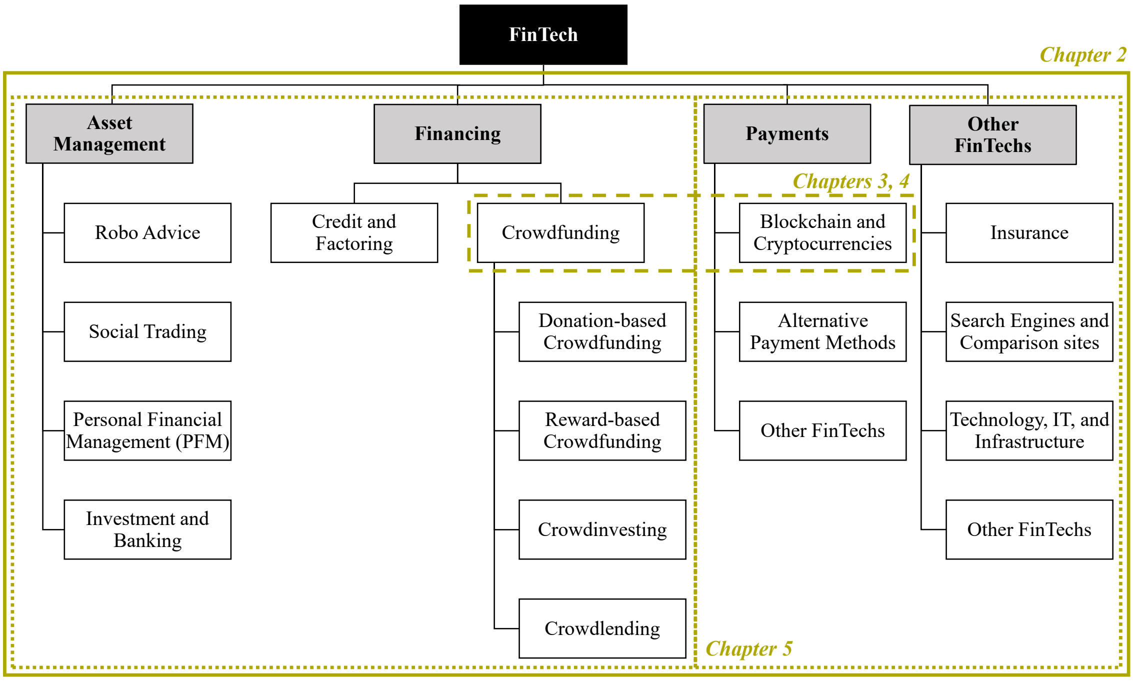


Figure 1.1: FinTech taxonomy of Dorfleitner et al. (2017) and attribution of FinTech (sub-)segments to the chapters of this dissertation. Own illustration based on Dorfleitner et al. (2017), p. 7.

The emergence of FinTech has garnered considerable attention in the aftermath of the global financial crisis of 2008, which eroded trust in traditional financial institutions (Goldstein et al., 2019). The growing interest in FinTech since then is reflected in the remarkable growth in global FinTech investments, soaring from $9 billion in 2010 to a peak of $239 billion in 2021, followed by a decline to $164 billion in 2022, attributable primarily to geopolitical and macroeconomic uncertainties and depressed valuations (KPMG International, 2023). Nevertheless, digitization had already gained traction in the financial sector before the financial crisis, exemplified by the proliferation of ATMs or mobile payments, encompassed by the broader terms e-finance or digital finance. In contrast, FinTech is a narrower concept focusing on purely technology-based innovation and process disruption (Gomber et al., 2017, 2018). This distinction stems from the fact that many of the new entrants to the FinTech industry initially operated in information technology, blurring the boundaries between finance and technology (Hendershott et al., 2021).

[1] Internet of Things is not directly considered in the taxonomy of Dorfleitner et al. (2017) and includes technologies to collect data on smart devices, which can be found in the sub-segments personal financial management, alternative payment methods, insurance, and technology, IT, and infrastructure.

In order to fully comprehend the FinTech market and its dynamics, academic research must also be devoted to this field. Empirical research on FinTech exhibits substantial differences compared to empirical research on traditional capital markets (Goldstein et al., 2019). The data basis of traditional empirical capital market research is often multi-year stock and standardized company data from licensed databases. Many FinTechs are young and relatively small companies, not listed on stock exchanges, and not subject to disclosure requirements. Thus, FinTech research lacks a common database. Consequently, new and creative approaches are required to obtain pertinent data to address research questions adequately. This data is frequently collected by hand or crawled automatically from websites. Alternative data in FinTech encompass texts, images, voices, videos, or digital footprints generated during surfing online (Cong et al., 2021). For instance, researchers have utilized texts of patent applications to classify and value FinTech innovations (Chen et al., 2019), satellite images of parking lots to study price informativeness that can discipline managers (Zhu, 2019), or digital footprints to predict consumer default (Berg et al., 2020). Hence, in FinTech research, it is often necessary to reconcile different data sources to ensure accuracy and to verify the actual execution of transactions.

Moreover, adopting digital platforms and technologies presents unprecedented opportunities to access completely new data sets and explore their potential applications (Feldman et al., 2015). The methods must be tailored to the underlying data, often requiring modifications to established approaches to capture the FinTech landscape's particularities effectively. Equally, established theories and explanatory approaches of traditional academic finance research can be applied to the new market and tested for validity. With this new data, research questions about financial technology can be addressed on two levels. The macro level initially scrutinizes all FinTechs and thus encompasses the entire industry. The micro level delves into the underlying business model or technological foundation of a FinTech on a more granular level, such as for individual projects.

Scientific publications on financial technology focused first on the micro level by investigating single sub-segments to analyze success or default determinants and probabilities in crowdfunding and peer-to-peer lending (Mollick, 2014; Dorfleitner et al., 2016), signaling in equity crowdfunding (Ahlers et al., 2015), learning effects in reward-based crowdfunding (Chemla and Tinn, 2020), FinTech competition in payment services for payment flows (Parlour et al., 2022), and to examine trading behavior and behavioral biases in social trading and robo advice (D'Acunto et al., 2019; Glaser and Risius, 2018). More recent studies emphasize the macro level of the entire industry to gain a holistic picture of digitization in the whole sector. Some studies consider location-related aspects (Cumming and Schwienbacher, 2018; Haddad and Hornuf, 2019), while others analyze whether FinTech innovation generates value (Chen et al., 2019), and FinTech access to finance (Bollaert et al., 2021). Another increasingly growing strand of literature links the societal discourse on sustainability to financial technology, owing to numerous potential synergies. Vismara (2019) investigates sustainability in equity crowdfunding, Merello et al. (2022) explore whether the sustainability profile of FinTechs is a value driver, and Tao et al. (2022) examine FinTechs as potential enablers of a low-carbon economy.

In the FinTech industry, there are very heterogeneous players, including young startups, established global financial services providers, and BigTechs, all of which are subject to different regulations. While traditional financial companies harbored concerns about competition from FinTechs, such reservations have dissipated, as collaborations can be mutually beneficial. FinTechs gain access to valuable experience or a banking license, while financial institutions augment their product portfolios with modern technologies (Klus et al., 2019). As a result, it is no longer possible to draw a clear line between FinTech and traditional financial institutions, and both worlds are becoming increasingly intertwined (Li et al., 2020). In addition, large technology companies, so-called BigTechs, are becoming significant players in the financial industry. They leverage their existing platforms and services to offer payment services, loans, insurance, and other financial products while capitalizing on synergies and network effects (Allen et al., 2021; Frost et al., 2020).

The transformation of the financial industry presents opportunities and threats to various stakeholders, such as investors, firms, and regulatory authorities. Initially, these new technologies improve access to finance for investors and companies (Bollaert et al., 2021). FinTechs are emerging as enablers of financial inclusion by extending mobile financial services to previously unbanked people and offering loans to new customers underserved by traditional banks (Allen et al., 2021; Erel and Liebersohn, 2022). As such, FinTechs contribute to achieving the Sustainable Development Goals, particularly in reducing inequality (Demir et al., 2022). Likewise, financially affluent users are drawn to FinTechs for their accelerated services and increased transparency (Allen et al., 2021). On the contrary, regulators express concerns about the potential threats to overall financial stability posed by FinTechs. As FinTechs increasingly collaborate with financial institutions, there is a risk of chain reaction effects that may contribute to systemic risk (Li et al., 2020). Furthermore, technology-based business models can cause money laundering, cyber security incidents, and data privacy risks, further raising regulatory concerns (Gomber et al., 2017, 2018).

One key aspect of digitization is that large amounts of data are generated and processed daily. This phenomenon has paved the way for innovative data-driven business models, particularly in the financial industry with FinTech firms. The *privacy calculus* model, derived from the literature on information economics, outlines the theoretical considerations of individuals about the disclosure of their personal data. Individuals conduct a cost-benefit analysis to weigh the costs and benefits of disclosing data (Dinev and Hart, 2006). Utilizing personal data can yield benefits for individuals, such as personalized advertising, as a foundation for AI applications, or for societal relevant issues (Dobkin, 2018). However, the misuse of the data can also result in negative consequences, including discrimination regarding economic or social aspects, not perceived influence or manipulation, censorship, and finally, threats to citizens' autonomy (Acquisti et al., 2015; Cohen, 2000; Dobkin, 2018). At least in the short term, the benefits of disclosing personal data often outweigh the costs, as potential risks are difficult to assess and will arise at some time in the future (Acquisti, 2004). Moreover, many users are often unaware that their data is being processed, further exacerbating these downsides (Acquisti et al., 2016).

The data processed in the financial industry, particularly by FinTechs, is sensitive and can reveal much about an individual's racial or ethnic origin, financial situation, political opinion, health, purchasing habits, location, and more (Dorfleitner et al., 2023). Consequently, policy intervention is required to balance the unequal distribution of interests, costs, and benefits between users and data-processing entities and to ensure transparency (Acquisti et al., 2015). On top of that, Acquisti et al. (2020) show that sufficient data privacy cannot be ensured by individuals alone, so it is imperative to have adequate regulation in place.

Processes in information technology and related data processing transcend national borders and occur under different jurisdictions due to the global proliferation of servers and cloud solutions. In response, the General Data Protection Regulation (GDPR) was implemented on May 25, 2018, in the European Economic Area as a comprehensive set of rules for cross-border data protection applicable to any processing of European personal data. The highest fine to date under the GDPR, amounting to €1.2 billion, was imposed on the company Meta in May 2023 for transferring European Facebook users' data to US servers.[2] From an academic research perspective, the GDPR represents a compelling natural experiment for examining how data-intensive companies, such as FinTechs, put data privacy into practice and to study the impact of the regulation in a before and after GDPR setting. A suitable and commonly used alternative data source for privacy analysis is the text of privacy statements which companies typically use to communicate their privacy practices and promote transparency to their users (Martin et al., 2017).

One of the main technological innovations driving FinTech adoption is the application of Distributed Ledger Technology (DLT) and its sub-type, the blockchain. The first and pioneering use of blockchain was the introduction of Bitcoin by Nakamoto (2008) as a decentralized peer-to-peer payment system (Hendershott et al., 2021). However, the potential application cases of blockchain in finance extend beyond its use as a means of payment, as it possesses versatile, unique technical features that disrupt existing market mechanisms and structures. These properties include a distributed database with peer-to-peer transmissions, transparency in conjunction with pseudonymity, and consistent, irreversible, tamper-proof entries (Cong and He, 2019; Gomber et al., 2018; Hendershott

[2] Murphy, H., Espinoza, J., 2023, May. Facebook owner Meta hit with record €1.2bn fine over EU-US data transfers. Financial Times.

et al., 2021; Tapscott and Tapscott, 2017; Yermack, 2017).

The growing academic literature in finance pertaining to blockchain technology covers several distinct sub-streams. One such sub-stream theoretically examines cryptocurrencies in economic and equilibrium models (Biais et al., 2023; Hinzen et al., 2022; Saleh, 2021), while another sub-stream explores valuation and alternative factor models for cryptocurrencies (Bianchi and Babiak, 2022; Liu and Tsyvinski, 2021; Liu et al., 2022; Makarov and Schoar, 2020). Likewise, other researchers compare cryptocurrencies to different asset classes and macroeconomics (Bianchi, 2020; Jiang et al., 2023; Yermack, 2015). The last sub-stream contains a plethora of studies that exhibit fundraising through initial coin offerings (ICOs) or security token offerings (STOs), with blockchain serving as the critical underlying technology (Fisch, 2019; Florysiak and Schandlbauer, 2022; Howell et al., 2020; Lyandres et al., 2022; Lambert et al., 2022; Thewissen et al., 2022).

Since practical implementation and proof of concepts besides Bitcoin are the most advanced in asset management, particularly blockchain-based tokens, this is a promising field for empirical research. A distinction is usually made at the legal level between payment tokens (such as Bitcoin), utility tokens, and security tokens. Utility tokens are issued through ICOs and used to fund projects, often representing vouchers for future services (Fisch, 2019; Howell et al., 2020). Since utility tokens lacked investor protection and many fraudulent cases occurred, the sentiment has deteriorated (Momtaz et al., 2019), and the development of security tokens was advanced. A security token digitally represents an investment product on the blockchain and typically falls under securities regulation (Lambert et al., 2022). As such, security tokens convey cash flow and potential ownership rights, including equity or debt, and represent a claim on the issuer's revenue rather than on future services as in the case of utility tokens (Sockin and Xiong, 2023).[3]

These blockchain-based forms of fundraising through ICOs and STOs represent a novel mechanism for companies to raise capital. The underlying intuition behind ICOs and STOs resembles crowdfunding, where many small investments from individual investors are aggregated by issuing multiple tokens that collectively reach the target amount. The utilization of blockchain technology automates the entire funding process and, unlike crowdfunding, enables the liquidation of shares through trading on secondary marketplaces (Lee et al., 2022). Consequently, these developments could disrupt the entire securities market structure, necessitating an analysis of how token markets operate and what determines

[3] In Germany, the enactment of the German Electronic Securities Act (eWpG) paves the way for electronic securities. It renders physical security certificates obsolete by mandating an entry in an electronic security register (§2(1) eWpG). Electronic securities can either take the form of bearer bonds (§1 eWpG) or investment share certificates (§95(1) KAGB). The subgroup of electronic securities also includes crypto securities (§4(3) eWpG) and crypto funds (KryptoFAV). Under the eWpG, they require an entry in a crypto security register (§16 eWpG). Hence, only some security tokens meet the definition of a crypto security or a crypto fund according to the eWpG (BaFin, 2023). In April 2023, a publicly disclosed draft bill revealed plans for extending the eWpG to shares.

STO market outcomes.

The theoretical framework of the well-known signaling theory can be applied to this nascent market. It can be used to develop an understanding of the behavior of the different market participants and how signals can overcome the potentially high information asymmetries between them (Spence, 2002). It is crucial to distinguish between the primary market, encompassing the STO issuance, and the secondary market, which entails trading tokens on various centralized and decentralized exchanges. The entire STO market can be examined by employing metrics commonly known for analyzing traditional securities offerings. These include indicators of STO success, underpricing, buy-and-hold (abnormal) returns, liquidity, and their associated determinants, with the corresponding adaptations to this new market.

Digital tokens can digitally represent and tokenize real-world assets on the blockchain, including real estate, commodities, fine art, and luxury watches. These tokenized assets are linked to and depend on the underlying asset outside the blockchain (Benedetti and Rodríguez-Garnica, 2023). Investors can benefit from various advantages offered by these tokens. They profit from lower entry barriers, can diversify their portfolios with modest amounts of money more broadly due to fractional ownership, and divest quickly, thereby enhancing liquidity. These benefits are often accompanied by lower transaction costs and times, resulting from the increasing automation of processes and related elimination of financial intermediaries (Lambert et al., 2022; Momtaz, 2023; Yermack, 2017).

Accordingly, it is of additional interest to delve deeper into a specific and alternative asset class. Since a well-diversified portfolio should contain real estate and the market for real estate tokens is developed, this asset class is suitable to study independently. The first question is whether such tokens deliver on their promise of portfolio diversification (Swinkels, 2023). Additionally, real estate tokens combine various aspects of real estate, crowdfunding, the crypto market, and the macroeconomy. These factors should be considered simultaneously through commonly-known fundamental factors of the underlying asset class, the sentiment and transaction costs in the crypto market, characteristics documented in crowdfunding, and the general macroeconomic situation to examine what is driving real estate token investors. Because many of these tokens are issued on the public and permissionless Ethereum blockchain, the underlying transactions are fully transparent, providing a novel data source for empirical research.

Consequently, the objective of this doctoral thesis is to make a contribution to the outlined areas of study through the following four distinct research papers, each co-authored with several collaborators. Table 1.1 provides an overview of these papers, the current publication status, and the assignment to the following chapters.

The remainder of this doctoral thesis is structured as follows. In the rest of the Introduction, the research papers are briefly summarized with respect to motivation, research questions,

Table 1.1: Overview of the publications with the corresponding chapter, title, and current publication status.

Chapter	Title	Publication	
		Status	Journal
2	Promise not fulfilled: FinTech, data privacy, and the GDPR	Accepted	Electronic Markets
3	Signaling in the Market for Security Tokens	Under review (Minor revision)	Journal of Business Economics
4	Real Estate Security Token Offerings and the Secondary Market: Driven by Crypto Hype or Fundamentals?	Conditional Acceptance	Journal of Banking & Finance
5	German FinTech companies: A market overview and volume estimates	Accepted	Credit and Capital Markets

data and method, and major results. Chapters 2 to 5 constitute the core of this doctoral thesis and present the four independent research papers. Chapter 6 concludes.

Promise not fulfilled: FinTech, data privacy, and the GDPR

Individuals intentionally and unintentionally disclose personal information online and when using their smartphones daily (Lindgreen, 2018; World Bank, 2021). Particularly FinTech firms process sensitive data that reveal much about an individual (Dorfleitner et al., 2023). Consequently, the first article of this dissertation sheds light on how the General Data Protection Regulation has affected the privacy practices of German FinTech companies. In doing so, it contributes to the literature on data privacy, particularly studies on the impact of privacy regulation and FinTech. Besides the theoretical analysis of Gai et al. (2017) and Ingram Bogusz (2018), surveys of Stewart and Jürjens (2018), and preliminary descriptives of Dorfleitner and Hornuf (2019), this research paper is to the best of our knowledge, the first to investigate data privacy of FinTech firms empirically.

Based on Figure 1.1, this paper studies the entire universe of FinTech companies in all (sub-)segments since every company has to deal with data privacy and privacy regulation. In the course of this, the analysis is embedded in the GDPR's guiding transparency principle for processing personal data (art. 5(1)a GDPR).

The data basis for this paper is the privacy statements of 276 German FinTech companies enriched with various company-and industry-specific variables. Our methodological approach is twofold, with textual analysis followed by multivariate analysis. The texts of the privacy statements are processed with standard methods in textual analysis. For readability, we compute the Neue Wiener Sachtext formula (Bamberger and Vanecek, 1984), SMOG metric adapted to the German language (McLaughlin, 1969; Bamberger and Vanecek, 1984), and alternatively a word count. Our metrics for standardization encompass cosine and jaccard similarity, as well as euclidean and manhattan distance. We construct indices for the quantity of data processed and the level of transparency. The indices are scaled in the interval between zero and one to enable the estimation of fractional probit regressions using quasi-maximum likelihood (Papke and Wooldridge, 1996). We have all the data before (pre) and after (post) the GDPR became binding. Thus, we can perform analysis based on t-tests and illustrations of the cumulative distribution functions. Furthermore, we calculate seemingly unrelated estimations using the stacking method, followed by Wald chi-square tests, to adequately compare the pre-and post-GDPR regression coefficients (Weesie, 1999; Zellner, 1962).

We find that the readability of the privacy statements has decreased, and the texts have become longer and require users more time to read. We document more standardized or boilerplate text suggesting that the information content users can draw from the statements is much lower. FinTechs prioritize technical and legal language to safeguard themselves over user comprehension, which contradicts the transparency principle of the GDPR. When analyzing the quantity of data processed, we find a significant increase, and the level of transparency remains unchanged. The number of external investors, as an indicator of external pressure, and the legal capital, which we interpret as the founders *ex-ante* commitment, have a positive effect on both the quantity of data processed and the level of transparency before the GDPR. Both effects disappear in the post-GDPR setting, as the GDPR was an incentive for all companies to focus on data privacy compliance.

Overall, we question whether users can give informed consent to process their data if they cannot transparently capture the language and the content of the privacy statements and, ultimately, whether the GDPR has achieved its objectives and fulfills its promises. Clearly, it is essential to acknowledge that full compliance with the GDPR may prove elusive for companies, as they would potentially need to restrict their business activities or even lose a competitive edge. This is particularly relevant in the rapidly evolving and data-driven FinTech industry.

Signaling in the Market for Security Tokens

In the following two research papers, we contribute to the literature on the confluence of two FinTech sub-segments. Specifically, crowdfunding, as a funding mechanism, combined with the underlying technological basis of blockchain and cryptocurrencies, as displayed in Figure 1.1. We investigate STO market outcomes along the entire life cycle of a security token, starting with the issuance through the STO in the primary market (pre-STO phase) and the subsequent trading on the secondary market (post-STO phase). Thereby, this paper is theoretically embedded in signaling theory to overcome the potentially high information asymmetries between the STO issuer and potential primary or secondary market investors due to this entirely new financing mechanism.

We study two hand-collected, overlapping, but non-identical data sets comprising 138 STOs and 108 security tokens traded on the secondary market. For the pre-STO phase, we examine whether a pre-sale before the main funding and the announcement of transferability are associated with the funding success. We measure funding success as the total funding amount and, alternatively, the ratio of the funding amount to the funding target as the degree of target achievement and estimate tobit models since the dependent variables are censored at zero. In the post-STO phase, the determinants of underpricing are investigated, measured as the return of the STO token price to the first market price. To this end, we estimate linear regressions and Heckman selection models, as the sample may exhibit a selection bias from token issuers underpricing their tokens to increase the chances of listing. For the post-STO performance, we calculate buy-and-hold returns (BHR) as well as buy-and-hold-abnormal returns (BHAR) adjusted by a value-weighted market capitalization-based benchmark for several short-term periods (Fisch and Momtaz, 2020; Momtaz, 2021a). Since decentralized exchanges (DEX) provide a new method of liquidity provision through so-called liquidity pools, a differentiation between tokens traded on centralized (CEX) and decentralized exchanges is made. Additionally, tokens theoretically promise liquidity, hence, we calculate liquidity measures based on Amihud (2002) and Amihud et al. (2006) or Corwin and Schultz (2012) for low-frequency markets.

First, we find that a pre-sale to collect information from potential investors early and the announcement of token transferability respective future expected liquidity serve as positive quality signals for investors to overcome information asymmetries before the STO. Second, we document hardly any underpricing on the secondary market. It is positively associated with the crypto market sentiment as an external signal, which remains robust in the Heckman model. Regarding the analysis of BHR and BHAR, we find both highly positive and negative returns, where a naïve diversification strategy is more promising to achieve higher returns. This finding demonstrates the high investment risk associated with security tokens, for which investors are only partially compensated. Lastly, we find that the secondary market lacks liquidity. DEXs are compared to CEXs, less liquid while offering lower barriers to entry which could enhance the entire market situation in the future.

In summation, we conclude that the overall security token market lacks professionalism in asset valuation and selection throughout our observation period. Notably, many theories commonly known from traditional capital markets offer limited explanatory power in this context.

Real Estate Security Token Offerings and the Secondary Market: Driven by Crypto Hype or Fundamentals?

The third article of this dissertation contributes to the small number of academic publications on real asset tokenization. The alternative asset class of real estate is the most relevant application case to study in this regard. So far, it has been examined mainly in general terms or from a theoretical, financial, legal, or technological perspective (Baum, 2021; Gupta et al., 2020; Liu et al., 2020a; Konashevych, 2020; Markheim and Berentsen, 2021). To the best of our knowledge, apart from Swinkels (2023) with a small sample, this article is the first to study real estate tokens empirically. We exemplify investor behavior and the determinants of tokenized real estate, such as fundamental factors affecting value, investment offering characteristics, and the crypto market's distinct features.

We investigate the property, financial, and crowdfunding characteristics of all 173 projects on the US real estate token platform RealToken as of December 2021. The project data are enriched with the related 238,433 blockchain transactions. To analyze portfolio construction and diversification in terms of investor behavior, a multidimensional blockchain transaction analysis on the wallet-investor, property-token level, and from the buy- and sell side is performed. Moreover, we explore the success of these real estate STOs by investigating the determinants of funding time, the number of days until 95% of tokens have been transferred to investors, and speed, the mean investment amount funded per day, in linear and accelerated failure-time survival models. Additionally, we examine aggregated daily capital flows in a panel data setting over time to study crypto-market-specific and macroeconomic determinants. By combining these three different methodological approaches, the market for real estate tokens is mapped holistically.

The results of our study underpin that real estate token investors do not yet hold well-diversified real estate token portfolios. These tokens provide broad access to real estate ownership for many small investors. Further, we conclude that investors acquire tokens mainly during the STO, and the secondary market plays a subordinate role. The property-specific fundamentals, crypto market-related transaction costs, and financial characteristics are positively related to STO success. Investors seek diversification possibilities through location choice to reduce the idiosyncratic cash-flow risk of the investment. Moreover, investors try to avoid high transaction costs not to reduce the return on investment. Interestingly, crowdfunding features are not associated with STO success, possibly due to low information asymmetries for properties. From the aggregated perspective of capital flows per day, we find that real estate token investors pay similar attention to the crypto market-specific sentiment and transaction costs when purchasing tokens. Only the transaction costs that directly reduce the return on investment are relevant for token sales. Macroeconomic factors have little effect on capital flows in general during our observation period.

Overall, the results underscore the importance of considering the specific crypto market environment alongside the characteristics of the underlying asset class for real asset tokenization. In this regard, the key novelty of this paper is to examine blockchain transaction data as a new data source to study investor behavior and the use of the constructed variables in multivariate analysis.

German FinTech companies: A market overview and volume estimates

This paper is divided into two parts. First, it provides a hand-collected market overview of all FinTechs operating in Germany, and second, an application case with volume estimates for the financing and asset management segments until December 2021.

The first part presents an extensive market overview and the related data description of FinTech companies operating in Germany as of December 2021. It is based on the data set of 978 German FinTech companies that has been published in the Mendeley Data repository as Dorfleitner et al. (2022) and can be accessed at https://doi.org/10.17632/438ytjyzxk.2. The data set was substantially extended qualitatively and quantitatively as part of the financial support of the Deutsche Bundesbank of project no. 7208857, which also gave rise to this paper. The market overview covers all FinTech segments and sub-segments, as displayed in Figure 1.1. The data collection process is outlined, followed by a description of the variables and an overview of the previous use of the data set. The descriptive statistics reveal that most FinTechs operate in the payments segment, followed by the sub-segment technology, IT, and infrastructure.

The second part of this paper provides an application case employing the data set for market volume estimates for the year 2021, drawing upon the estimation techniques outlined in Dorfleitner et al. (2017) and Dorfleitner et al. (2020) along with the previous years' estimations in Dorfleitner and Hornuf (2023). The volume estimates specifically pertain to the asset management and financing segments and encompass all their sub-segments since the volumes are only meaningful and available for these segments (refer to Figure 1.1). In the financing segment, which contains the crowd-based forms of funding, market volumes correspond to real *transaction volume*, representing the money raised. Market volumes in the asset management segment are comparable to assets under management, concretely invested money. Since 2015, the FinTech market has exhibited steady, linear growth, with a peak of €85.3 billion in 2021 and growing at a rate of 28 percent compared to 2020.

In conclusion, this paper substantially contributes to comprehending the entire German FinTech market while providing valuable data and information for researchers, practitioners, supervisory authorities, and regulators in this dynamic industry.

Chapter 2

Promise not fulfilled: FinTech, data privacy, and the GDPR

This research project is joint work with Gregor Dorfleitner (University of Regensburg) and Lars Hornuf (TU Dresden). The paper has been published as:

Dorfleitner, G., Hornuf, L., Kreppmeier, J. (2023). Promise not fulfilled: FinTech, data privacy, and the GDPR. *Electronic Markets*, forthcoming.

Abstract This article analyzes how the General Data Protection Regulation (GDPR) has affected the privacy practices of FinTech firms. We study the content of 276 privacy statements respectively before and after the GDPR became binding. Using text analysis methods, we find that the readability of the privacy statements has decreased. The texts of privacy statements have become longer and use more standardized language, resulting in worse user comprehension. This calls into question whether the GDPR has achieved its original goal—the protection of natural persons regarding the transparent processing of personal data. We also link the content of the privacy statements to FinTech-specific determinants. Before the GDPR became binding, more external investors and a higher legal capital were related to a higher quantity of data processed and more transparency, but not thereafter. Finally, we document mimicking behavior among FinTech industry peers with regard to the data processed and transparency.

Keywords Data privacy, FinTech, General Data Protection Regulation, Privacy statement, Textual analysis, Financial technology

JEL K200, L810, M13

2.1 Introduction

Data have become a critical resource for many business models as a result of digitalization and globalization. Individuals disclose personal information intentionally and unintentionally over the Internet and when using their smartphones (Lindgreen, 2018; World Bank, 2021). Because of the international location of servers and cloud-computing services, the processing of data often takes place under different jurisdictions and does not stop at national borders. On May 25, 2018, the General Data Protection Regulation (GDPR) became binding in the European Economic Area (EEA)[1] to address the increasing challenges of data security and privacy. The GDPR extends its territorial reach even outside the EEA if European data are involved. The financial sector and, in particular, the recently emerging Financial Technology (FinTech) industry process large amounts of sensitive data. Payment data, for example, can entail information about racial or ethnic origin, political opinions, religious beliefs, trade-union membership, health or sex life. The different FinTech business models, which frequently rely on artificial intelligence, big data, and cloud computing, thus represent an important and relevant industry to examine the impact of the GDPR on data privacy practices.

Companies are not required by law to have a privacy statement; however, they often comply with the requirement to inform their users (art. 13-15 GDPR), by publishing such statements, about the personal data they process. Therefore, privacy statements serve as research objects for many studies that analyze privacy. For example, Ramadorai et al. (2021) study a signalling model of firms engaging in data extraction. They analyze a sample of 4,078 privacy statements of U.S. firms and find significant differences in accessibility, length, readability and quality between and within the same industries. Large companies with a medium level of technical sophistication appear to use more legally secure privacy statements and are more likely to share user data with third parties. Other studies analyze the effect of privacy regulation by comparing privacy-statement versions before and after the GDPR became binding. Becher and Benoliel (2021), for instance, focus on the "clear and plain language" requirement in the GDPR (art. 12 GDPR). By analyzing the readability of 216 privacy statements of the most popular websites in the United Kingdom and Ireland after the GDPR became binding, they conclude that privacy statements are hardly readable. For a small sub-sample of 24 privacy statements before and after the GDPR became binding, they document a small improvement in readability. In another study, Degeling et al. (2019) periodically examine, from December 2017 to October 2018, the 500 most popular websites of all EU member states, gathering a final sample of 6,579 privacy statements, and find that the number of sites with privacy statements increased after the GDPR became binding. When focusing on cookie consent libraries,

[1]Thus, it applies in the European Union (EU) and the three countries of the European Free Trade Association except Switzerland.

they conclude that most cookies do not fulfill the legal requirements. Linden et al. (2020) study 6,278 privacy statements inside and outside the EU. They underline that the GDPR was a main driver of textual adjustments and that many privacy statements are not yet fully compliant regarding disclosure and transparency. This article extends the previous research by focusing on the FinTech industry in its entirety, which is characterized by the presence of companies in different growth stages ranging from startup companies to established global corporations. Data privacy is particularly important for FinTechs who find themselves caught between the pressure to innovate for future business success and the privacy aspects that result from the highly sensitive data processed in financial services. To address the peculiarities of the companies within the FinTech industry and the data they process, we link the analysis of privacy statements to company- and industry-specific factors.

The guiding principle of the processing of personal data according to the GDPR is transparency (art. 5(1)a GDPR). In this paper, we analyze 276 privacy statements published by German FinTech firms before and after the GDPR became binding. We analyze the readability of the privacy statements, their standardization as a basic requirement for transparency, the amount of data processed, and transparency of data processing in the true sense. We then examine how FinTech company and industry specific factors influence these metrics. We perform textual analysis on the privacy statements and provide evidence that their readability has worsened since the GDPR became binding. Specifically, the texts have become longer and more time-consuming to read. In a next step, we find an increase in the use of standardized text. Further, we study the quantity of data processed as stated in the privacy statements and the related level of transparency. We study whether FinTech-specific factors such as the number of external investors and the existence of bank cooperation predict privacy practices respectively before and after the GDPR became binding. Finally, peer pressure among FinTechs and industry standards might induce mimicking behaviour. We find that *ex-ante* industry-wide privacy practices influence FinTechs' privacy practices after the GDPR became binding. Our results remain robust when excluding more mature FinTechs and when using alternative model specifications.

The rest of this article is organized as follows. The "Institutional Background: The GDPR" section describes the institutional background of the GDPR and the theoretical framework of this study. The "Literature and Hypotheses" section examines the related literature and develops the hypotheses that will be tested. The "Data and Method" section outlines the data and method. The "Results" section presents our results. The "Robustness" section provides robustness checks, and the "Conclusion" section concludes.

2.2 Institutional Background: The GDPR

The European Parliament passed the GDPR on April 14, 2016. After a transition period, the regulation became binding on May 25, 2018. The regulation is intended to harmonise data protection legislation in the EU. According to its territorial scope (art. 3 GDPR), data of EU citizens are subject to the regulation, independent of whether the data are processed inside or outside the EU. After the GDPR became binding, many jurisdictions outside the EU adopted data protection regulations with a scope and provisions similar to those in the GDPR.[2] In addition to questions of data security, the GDPR distinguishes between four main actors in the field of privacy: the data subject, who is a natural person and whose personal data are processed; the data controller, as the entity offering products or services for which the data are needed; the data processor, supporting the data controller to process the data; and third parties that might process data not directly related to the product or service provision (e.g., companies evaluating a user's credit-worthiness) (Linden et al., 2020). To give the GDPR bite, fines of up to 4% of a company's yearly global revenue or 20 million euros can be imposed in cases of non-compliance (art. 83 GDPR).

This article builds on art. 5 GDPR, which describes the key principles of the processing of personal data, in particular the overarching principle of transparency.[3] Art. 5 GDPR is further specified in the rec. 39 GDPR which demands *inter alia* that natural persons should transparently know about the form of processing of their personal data and the extent of data processing. The basic requirement for transparency is that the information is communicated in understandable language.[4] In addition, our analysis is based on the more concertising statements by the Article 29 Article 29 Working Party (2018). Based on the aforementioned legislation regarding transparency, we further investigate in this study the theoretical concepts of readability, standardization, quantity of data processed and finally transparency which we subsume under the term privacy practices.

An important EU directive that pertains directly to the GDPR and which deals with data protection in the FinTech sector—especially in payment services—is the Payment Services Directive 2 (PSD2). Focusing on payment services, the PSD2 regulates practices related to the processing of payment data and lawful grounds for granting access to bank accounts. The PSD2 also deals with the processing of silent party data. Silent party data is personal data of a data subject who is not a user of a specific payment service provider,

[2] Specific examples of privacy regulations similar to the GDPR are the California Consumer Privacy Act of 2018, the Personal Data Protection Act 2019 in Thailand, the Brazilian General Data Protection Law of 2020, the Swiss Federal Act on Data Protection of 2020, and the Chinese Personal Information Protection Law of 2021.

[3] "Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject" (art. 5(1)a GDPR).

[4] "The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used." (rec. 39 GDPR).

but whose personal data is processed by that payment service provider for the performance of a contract between the provider and a payment service user. Similar to the GDPR, the PSD2 also addresses issues of user consent, data minimization, data security, data transparency, data processor accountability, and user profiling. Although the PSD2 affects some of the FinTechs studied in this article, we focus below on the more general GDPR, which is equally applicable to all FinTechs.

2.3 Literature and Hypotheses

2.3.1 Related literature

The theoretical foundation of this study is embedded in the economics of privacy literature investigating economic trade-offs that reveal people's considerations in terms of privacy.[5] The economics of privacy literature is embedded in the broader context of information economics (Posner, 1981) and is substantially affected by the advances in digital information technology.

The GDPR as a new data protection regulation affects nearly every area of life where natural persons claim a service or product with or in exchange for personal data. Therefore, the encompassing consequences and the economic impact of the GDPR are quantified in several studies and highlight a decrease in web traffic, page views and revenue generated as a result of the consent requirement on the part of the data subject (art. 7 GDPR) or limitations in marketing channels (Aridor et al., 2020; Goldberg et al., 2021).

Privacy statements are the essential source of information about how companies put privacy into practice and process personal data. These statements are the standard way to promote transparency to users (Martin et al., 2017) and to balance the equity of power between data subjects and data processors (Acquisti et al., 2015). Therefore, privacy statements are often used in the literature to analyze privacy-related aspects of companies as outlined in the Introduction. Computer and information science scholars have developed tools that help researchers analyze privacy statements on a large scale (Contissa et al., 2018; Harkous et al., 2018; Tesfay et al., 2018). Contissa et al. (2018), for example, apply their tool to the privacy statements of large-platform and BigTech companies as an exploratory inquiry and conclude that none fully comply with the GDPR, as the formulations are partially unclear, potentially illegal or insufficiently informative.

Privacy and security aspects of FinTech companies have been studied in a variety of contexts. Stewart and Jürjens (2018) survey the German population regarding FinTech adoption and identify data security, consumer trust and user-design interface as the most important

[5]For a literature review on the economics of privacy, see Acquisti et al. (2016).

determinants. Gai et al. (2017) provide a theoretical construct for future FinTech industry development to ensure sound security mechanisms based on observed security and privacy concerns and their solutions. Other studies emphasize the specificity and importance of the data processed by FinTechs. Ingram Bogusz (2018) describes and distinguishes the data that FinTechs process between *content data*, directly related to the identification of a person, and *metadata*, usually left unintentionally by users but useful for the data processor. Berg et al. (2020) demonstrate the large opportunities to use data collected during 250,000 purchases on a German e-commerce website. Among other things, such data has significant explanatory power to determine creditworthiness. Dorfleitner and Hornuf (2019) provide a descriptive analysis of privacy statements of German FinTechs before and after the GDPR became binding to derive policy recommendations. However, apart from Dorfleitner and Hornuf (2019), the preliminary research does not analyze the privacy statements of FinTech companies specifically regarding privacy regulation and the GDPR. In this study, we go well beyond the simple descriptive statistics of Dorfleitner and Hornuf (2019) and examine the readability and standardization of privacy statements using text analysis. Furthermore, we link the content of the FinTechs' privacy statements to company- and industry-specific factors in a multivariate context in order to account for the diversity and specificity of business models within the FinTech industry.

2.3.2 Derivation of Hypotheses

Readability The GDPR requires that information and communication be transmitted to users in clear and plain language (art. 5, 7, 12 GDPR, rec. 39, 42, 58 GDPR) in order to achieve transparency. This objective corresponds to the linguistic concept of readability, i.e. the reader's ease with and ability to understand a text. Apart from the legislative requirements of the GDPR, companies also have an economic incentive to provide readable privacy statements, which in turn can increase user trust in their business conduct (Ermakova et al., 2014) and thereby create a competitive advantage (Zhang et al., 2020). While these arguments seem to suggest that companies should have increased the readability of their privacy statements after the GDPR became binding, there are also severe counterarguments. Many users do not read disclosures such as privacy statements (Omri and Schneider, 2014), even for products and services they use daily (Strahilevitz and Kugler, 2016). Firms provided their users, often within a very short time frame, updated privacy statements after the GDPR became binding (Becher and Benoliel, 2021). It appears unlikely that such a large number of new privacy statements has triggered additional engagement with these texts by data subjects. Indeed, several studies state that privacy statements are difficult and time-consuming to read and often require an understanding of complex legal or technical vocabulary (Fabian et al., 2017; Lewis et al., 2008; Sunyaev et al., 2015). Second, and in line with this observation, Earp et al. (2005)

and Fernback and Papacharissi (2007) find that privacy statements often aim to protect companies from contingent lawsuits rather than address the privacy needs of data subjects. Thus, while firms know that their customers tend to ignore privacy statements, especially if they are technical to read, they may have emphasized their own interests with respect to avoiding lawsuits when updating these statements with respect to the GDPR. Indeed, as long as there is no need for companies to fear that the requirement of clear and plain language will become the subject of legal proceedings, they have few incentives to improve the readability of their privacy statements.

This theoretical argumentation is supported by empirical evidence. Two years after the GDPR became binding, the penalties imposed on companies remain relatively low, and none traces back to the clear and plain language requirement (Wolff and Atallah, 2021). For a sample of 24 privacy statements from the most popular websites in the United Kingdom and Ireland, Becher and Benoliel (2021) finds that many of the privacy statements before the GDPR were barely readable and have improved only slightly since the GDPR became binding. Linden et al. (2020) study 6,278 privacy statements before and after the GDPR became binding using different text metrics like syllables, word count or passive voice and state that the policies became significantly longer but that there was no change in sentence structure.

Summarizing this reasoning, we expect that companies may not have significantly improved the readability of their privacy statements after the GDPR became binding in May 2018.

> **Hypothesis 1:** *The readability of FinTech privacy statements has not improved since the GDPR became binding.*

Standardization The standardization of legal text is often deemed uninformative for the reader and is therefore referred to as *boilerplate* in academic literature. Boilerplate language is characterized by very similar uses of language and wording across legal documents from different issuers (Peacock et al., 2019) and little company-specific information (Brown and Tucker, 2011). For a user, boilerplate text requires much effort to read, and details might appear to be irrelevant (Bakos et al., 2014).

Boilerplate language in legal text brings cost advantages for companies. First, the costs of adopting the specific legal requirements such as the GDPR are lower for all market participants. Second, reduced legal uncertainty due to the use of established and proven text passages, which have yet to cause legal violations, promises fewer future penalties (Kahan and Klausner, 1997). For many companies, the GDPR provided an incentive to intensively address and spend resources on data privacy compliance (Martin et al., 2019). During the period of transition to the GDPR, organizations looked for external information and support regarding the implementation of its legal requirements. Companies often rely on compliance assessment tools to audit their business processes for legal compliance (Agarwal et al., 2018; Biasiotti et al., 2008). In the related literature of requirements engineering, boilerplate language is often proposed to reduce text ambiguities (Arora et al., 2014). For example, Agarwal et al. (2018) provide a tool specifically designed for assessing GDPR compliance, including one process step that allows the user to incorporate boilerplate language. Other sources of information are websites or online policy generators, which deliver guidance on implementing and interpreting the GDPR or even templates for generating privacy statements.[6] The mentioned advantages of applying boilerplate language as well as the examples of assistance to GDPR compliance underpin that we can expect an increase in boilerplate language in the privacy statements since the GDPR became binding.

> **Hypothesis 2:** *The standardization of FinTech privacy statements has increased since the GDPR became binding.*

Quantity of data processed and transparency For a comprehensive analysis of the FinTechs' transparency beyond readability and standardization, we investigate the content of the privacy statements. While the mere quantity of data processed is important in a first step, we also consider the actual level of transparency.

At the core of the GDPR are principles related to the processing of personal data (art. 5 GDPR), in particular the articles related to lawful, fair and transparent data processing as well as data minimization (art. 5 (1a, c), rec. 39 GDPR). An increase in transparency ensures that consumers provide better-informed consent with respect to the data processed (art. 4, 11 GDPR) (Betzing et al., 2020). An imprecise statement about which and how much personal data are processed violates the provisions of the GDPR, which in turn can result in high penalties. Thus, with regard to the expected costs, an accurate disclosure about which data are processed outweighs the general principle of data minimization. However, the major change of the GDPR introduced compared with the previous privacy legislation in Germany is the potential for high penalties (Martin et al., 2019). This fact represents an incentive for companies to rework their privacy statements, to be precise about the quantity of data processed and to enhance transparency after the GDPR became binding.

Regarding the behavior of data subjects, we apply the theoretical considerations of the privacy calculus model. Data disclosure is the result of a consumer's individual cost-benefit analysis, referred to as a privacy calculus, according to which costs and benefits of disclosing personal data are weighed against each other (Dinev and Hart, 2006). The potential risks of data disclosure are difficult to assess and will only appear in the future, which is why

[6] A template for privacy statements funded within the Horizon 2020 Framework Program of the European Union is provided at https://gdpr.eu/privacy-notice/, last access: 31 August 2021.

benefits often outweigh costs in the short run (Acquisti, 2004). Data subjects must consent to the privacy statements that are written by companies if they are to receive immediate gratification (O'Donoghue and Rabin, 2000) or, more concretely, to obtain a desired service or product (Aridor et al., 2020). The notion behind many business models is that customers actively forsake parts of their data privacy in exchange for goods and services (Mulder and Tudorica, 2019). Therefore, the data subject's control over the data processed and transparency is limited, and companies have the upper hand.

Empirical studies evidence that it is beneficial and important for companies to ensure and enhance transparency. Li et al. (2019) show that transparency may enhance trust and reputation in a business's activities. Martin et al. (2017) find that a higher level of transparency in the case of a data breach results in a lower negative stock-price reaction.

To summarize the argumentation, we expect an increase not only in the quantity of data processed but also in transparency as companies fulfill the legal requirements of the GDPR and avoid potentially high penalties while benefiting economically.

> **Hypothesis 3a:** *The quantity of data processed by FinTechs has increased since the GDPR became binding.*

> **Hypothesis 3b**: *The transparency of FinTechs has increased since the GDPR became binding.*

Determinants of both the quantity of data processed and transparency In order to account for the peculiarities and diversity of the FinTech industry with regard to data privacy practices, we pay particular attention to the finance literature in developing the following hypotheses. Young companies, such as most FinTechs, prioritize the core business instead of privacy compliance when launching a seminal business. Moreover, founders are rarely experts in privacy or law. Nevertheless, when starting business operations, FinTechs inevitably process personal data and need to act in order to protect privacy sufficiently (Miller and Tucker, 2009) and to comply with current privacy regulation. Therefore, the question arises whether some FinTechs meet the legal requirements better than others. External investors contribute knowledge and experience to build a proper and future-oriented company. The advanced knowledge of external investors is based on experience in legal compliance and privacy with corresponding business contacts and cooperations (Hsu, 2006). The more external investors are involved in an investment, the more likely it is to succeed as a business because of the access to external knowledge (De Clercq and Dimov, 2008). We hypothesize that having a greater number of investors with different education, experience and background knowledge help achieve privacy compliance.

> **Hypothesis 4a:** *External investors increase both the quantity of data processed and transparency of FinTechs.*

Another important group of stakeholders for FinTechs are the banks they may collaborate with. Within such cooperation, FinTechs receive access to financial resources, infrastructure, customers, security reputation (Drasch et al., 2018), a banking license and legal support to comply with regulation (Hornuf et al., 2021a). Moreover, banks have a strong incentive to collaborate with FinTechs in order to boost their digital transformation, which might result in more data being shared. Banks also have long-term experience managing personal data and handling data in compliant way. Banks can transfer this knowledge to FinTechs, especially if they cooperate. We therefore expect that cooperation with a bank has a positive effect on compliance with privacy regulation.

> **Hypothesis 4b:** *Cooperations with banks increase both the quantity of data processed and transparency of FinTechs.*

Mimicking behavior Mimicking behavior often leads to standardization (Kondra and Hinings, 1998) as described in Hypothesis 2, which is particularly likely to be at work after the GDPR became binding. Prior studies evidence that companies tend to mimic the behavior of other companies in the same industry, including for stock repurchase decisions (Cudd et al., 2006), target amounts in crowdfunding (Cumming et al., 2020) or tax avoidance (Kubick et al., 2015). An industry-centric perspective with regard to privacy appears to be reasonable; as Martin et al. (2017) show, when a specific entity experiences a privacy breach, the firm performance of companies in the same segment is also affected. In our study, FinTechs operating in the same sub-segment and thus having corresponding business models should also have similar data processing practices (Hartmann et al., 2016). Consequently, there is an incentive to adopt an immediate peer's privacy statement. Mimicking an industry peer's behavior in the field of privacy is fairly easy, as the privacy statements can be accessed on the corresponding website with just a few clicks. Firms in the same segment can expect to incur similar fines and penalties in cases of non-compliance (Hajduk, 2021). Expert interviews in the context of the GDPR reveal that start-up executives have concerns that their industry peers could report their possible violations to the data protection authorities (Martin et al., 2019). Mimicking industry peers and adopting similar privacy practices prevents companies from experiencing such adversity.

We therefore expect that the industry-specific design of privacy practices stated in privacy statements has a positive influence on a single company's quantity of processed data and transparency.

> **Hypothesis 5:** *Mimicking behavior has a positive influence on the company-specific quantity of data processed and transparency.*

2.4 Data and Method

2.4.1 Data

Our sample consists of companies operating in financial technology in Germany.[7] Data collection before the GDPR became binding, on 25 May 2018, took place between 15 October 2017 and 20 December 2017. Data collection after the GDPR became binding occurred between 15 August 2018 and 31 October 2018. We comprehensively map the FinTech industry operating in Germany and include both FinTech start-ups and established FinTech companies in our sample. The sample consists of 276 companies with German privacy statements.

2.4.2 Variables

To test Hypothesis 1, we use the readability measures *SMOG German, Wiener Sachtext* and, alternatively, *No. words*. For a test of Hypothesis 2 to examine standardization, we calculate the similarity and distance metrics *Cosine similarity, Jaccard similarity, Euclidean distance* and *Manhattan distance*. We describe these text-based measures and their respective calculations in more detail in the "Methods" section.

2.4.2.1 Variables of interest

To test Hypotheses 3a and 3b, 4a and 4b and 5, we construct a *data index* to account for the quantity of data processed and the *transparency index* for actions undertaken to ensure transparency. The underlying assumption of the index construction is that we assume that when a company does not concretely state the processing of specific data or certain data-processing practices, such processing does not occur. After the GDPR became binding, this assumption seems justified given the high potential penalties for misrepresentation.

The *data index* is a measure of the quantity of data processed by a company. The data processed ranges from general personal data (e.g. name, address) to metadata (e.g. IP address, social plugins) to special categories of personal data (e.g. health, religion). Table 2.1 provides the full list of data categories from which the *data index* is composed. For the variable *transparency index*, we aggregate variables representing different dimensions of transparent data-protection actions undertaken by the companies.

[7] Study data are kindly provided by Dorfleitner and Hornuf (2019). We reduced the original data set to 276 companies because of the non-availability of privacy statements, non-availability of privacy statements in German language, inconsistencies in company data and inactivity or insolvency during both data collection periods.

Apart from vague formulations in art. 12 and rec. 58, 60, the GDPR does not explicitly define and specify transparency or how to ensure transparency. Therefore, we combine the potential transparency vulnerabilities of Mohan et al. (2019) and Müller et al. (2019) to define our considered dimensions of transparency. The *transparency index* represents the normalized sum over eight dummy variables such as *data* (whether a company states in detail which personal data they process), *purpose* (1 if a company states for what reason or purpose personal data are processed) and *storage* (1 if it states how long data are stored or when they are deleted). Table 2.1 lists in detail the composition of the *transparency index.*

As proposed by Wooldridge (2002, p. 661), we divide the indices data index and *transparency index* by the maximum achievable number of variables of which the respective index is composed to scale them between 0 and 1. We interpret a higher index value to mean respectively a higher quantity of data processed and more transparency.

Table 2.1: Definition of variables.

Variable	**Description**	**Source**
Bankcooperation	D: 1 if the Fintech cooperates with a bank, 0 otherwise.	Bank, FinTech websites
No. investors	Logarithm plus 1 of the number of external investment firms and individual investors	BvD Dafne, Crunchbase
Mimic Data Index	Mimicking variable for Data Index	Own calculations
Mimic Transparency Index	Mimicking variable for Transparency Index	Own calculations
Wiener Sachtext	Neuer Wiener Sachtext readability metric	Own calculations
SMOG German	SMOG readability metric (adopted to German language)	Own calculations
No. words	Logarithm of the total number of words	Own calculations
Cosine similarity	Cosine similarity	Own calculations
Jaccard similarity	Jaccard similarity	Own calculations
Euclidean distance	Euclidean distance	Own calculations
Manhattan distance	Manhattan distance	Own calculations
Controls		
Firm age	Logarithm of the age of the FinTech company.	German company register, LinkedIn
Employees	Number of employees (rank variable between 1 and 5)	BvD Dafne, Crunchbase, LinkedIn
City	D: 1 located in a city with more than one million inhabitants, 0 otherwise.	German company register, Websites
Legal capital	D: 1 if a company has a legal form that requires a legal capital of more than 1 EUR, 0 otherwise.	German company register, Websites
GDPR	D: 1 if observations are after the introduction of the GDPR on May 25th 2018, representing the post-GDPR period, 0 otherwise.	
Data index	An index aggregating the quantity of data processed. The index adds the hereafter following variables and is divided by 38.	Own calculations
Name	D: 1 if the first and last name are processed, 0 otherwise.	D and H (2019)
Gender	D: 1 if the gender or form of address are processed,	D and H (2019)

Continued on Next Page

Variable	**Description**	**Source**
	0 otherwise.	
Title	D: 1 if the title is processed and 0 otherwise.	D and H (2019)
Language	D: 1 if the company processes the language, 0 otherwise.	D and H (2019)
Identifier	D: 1 if the identifier (e.g. user name or ID) is processed, 0 otherwise.	D and H (2019)
Password	D: 1 if the the password is processed, 0 otherwise.	D and H (2019)
Age	D: 1 if the age or date of birth are processed, 0 otherwise.	D and H (2019)
Place of birth	D: 1 if the place or country of birth are processed, 0 otherwise.	D and H (2019)
Address	D: 1 if the address or delivery address or billing address are, processed, 0 otherwise.	D and H (2019)
E-mail address	D: 1 if the e-mail address is processed, 0 otherwise.	D and H (2019)
Phone number	D: 1 if the phone number or mobile number are processed, 0 otherwise.	D and H (2019)
Residence city	D: 1 if the city of residence is processed, 0 otherwise.	D and H (2019)
Residence country	D: 1 if the company processes the country of residence, 0 otherwise.	D and H (2019)
Marital status	D: 1 if the company processes the marital status, 0 otherwise.	D and H (2019)
Occupation	D: 1 if the occupation or employee status are processed, 0 otherwise.	D and H (2019)
Bank	D: 1 if the bank data or account data or payment data are processed, 0 otherwise.	D and H (2019)
PIN	D: 1 if the PIN or TAN are processed, 0 otherwise.	D and H (2019)
Income	D: 1 if the monthly revenues or expenses are processed, 0 otherwise.	D and H (2019)
Tax residency	D: 1 if the tax residency or status are processed, 0 otherwise.	D and H (2019)
Social security number	D: 1 if the social security number is processed, 0 otherwise.	D and H (2019)
Tax ident number	D: 1 if the tax identification number is processed, 0 otherwise.	D and H (2019)
Driving license	D: 1 if driving license data is processed, 0 otherwise.	D and H (2019)
Passport, registration	D: 1 if passport and identity card data or the registration number are processed, 0 otherwise.	D and H (2019)
Graduation, qualification	D: 1 if information on graduation or qualifications are processed, 0 otherwise.	D and H (2019)
Insurance	D: 1 if information on insurance is processed, 0 otherwise.	D and H (2019)
IP-address	D: 1 if the IP-address is processed, 0 otherwise.	D and H (2019)
GPS, location	D: 1 if GPS or location data are processed, 0 otherwise.	D and H (2019)
Personal data published	D: 1 if personal data are published, 0 otherwise.	D and H (2019)
Personal data transfer	D: 1 if personal data are collected from, transferred to or disclosed with third parties, 0 otherwise.	D and H (2019)
Social Plugins, Third party	D: 1 if social plugins are used or third party services are integrated, 0 otherwise.	D and H (2019)
Behavior, usage, movement	D: 1 if behavioral, usage or movement data are processed or tracking services are used, 0 otherwise.	D and H (2019)
Google Analytics	D: 1 if Google Analytics is used, 0 otherwise.	D and H (2019)
Health	D: 1 if health-related data is processed, 0 otherwise.	D and H (2019)
Religion	D: 1 if the religious confession is processed, 0 otherwise.	D and H (2019)
Nationality	D: 1 if the nationality or citizenship is processed, 0 otherwise.	D and H (2019)
Picture	D: 1 if user or title pictures are processed, 0 otherwise.	D and H (2019)
Conversation record	D: 1 if a conversation recording is processed, 0 otherwise.	D and H (2019)

Continued on Next Page

Variable	**Description**	**Source**
Signature	D: 1 if the signature or sample of writing is processed, 0 otherwise.	D and H (2019)
Transparency index	An index aggregating dimensions of transparency we define hereafter. The index adds the hereafter following variables and is divided by 8.	Own calculations
Data	D: 1 if the company states which personal data are processed, 0 otherwise.	D and H (2019)
Purpose	D: 1 if the company states for what reason or purpose personal data are processed, 0 otherwise.	D and H (2019)
Storage	D: 1 if the company states for how long data are stored or when they are deleted, 0 otherwise.	D and H (2019)
Avoid	D: 1 if the company states if there exists a possibility to avoid data processing, 0 otherwise.	D and H (2019)
Opt-In	D: 1 if the company states whether they have an Opt-In procedure, 0 otherwise.	D and H (2019)
Pseudo	D: 1 if the company states that data are processed pseudonymously, 0 otherwise.	D and H (2019)
Third	D: 1 if the company states which personal data are shared with third parties, 0 otherwise.	D and H (2019)
Third data	D: 1 if the company states with which third parties data are shared, 0 otherwise.	D and H (2019)
		End of table

Note: List and definitions of all variables with the corresponding source. In the following table the abbreviation “D" stands for dummy variable and “D and H (2019)" for Dorfleitner and Hornuf (2019). All variables that are directly included in the following analyzes are marked in *italics*.

2.4.2.2 Explanatory variables

To construct our explanatory and control variables, we collect detailed firm-specific variables, which we describe below with the data sources used. Accuracy of the data was validated using cross-checks with press releases, FinTech websites and other news and information online.

To test Hypothesis 4a, that a higher number of external investors positively influences the quantity of data processed and transparency, we include the variable *No. investors*, measured as the absolute number of external investment firms and individual investors who funded the company. This variable is already considered in other FinTech-related studies such as Cumming and Schwienbacher (2018) and Hornuf et al. (2018b). We derive the variable from the BvD Dafne and Crunchbase database, which was also used in other academic papers, such as Bernstein et al. (2017) and Cumming et al. (2019).

To test Hypothesis 4b, we include the dummy variable *bankcooperation*, which equals 1 if the respective company has a cooperation with a bank and 0 otherwise. For data collection, we first searched all bank websites to find indications of bank–FinTech cooperation. In a second step, we checked for cooperation from the FinTech side.

To analyze mimicking behavior as outlined in Hypothesis 5, we follow the approach of

Cudd et al. (2006), who use the industry average of a measure in the year preceding the focal period for mimicking behavior. We obtain the variables *mimic data index* and *mimic transparency index* by calculating the average of the indices *data index* and *transparency index* within the same FinTech sub-segment before the GDPR became binding according to the taxonomy of Dorfleitner et al. (2017).

2.4.2.3 Control variables

To consider unobserved heterogeneity, we use the following control variables. First, we control for firm location with the variable *city*, which can be a relevant geographic determinant. This variable indicates whether a company is located in a city with more than one million inhabitants. In metropolitan areas, more customers and sources for funding (Hornuf et al., 2021a) as well as start-up incubators are within geographical reach and thus available to support a company's development. Besides, more FinTechs are located in one place in metropolitan areas, which often leads to the establishment of entrepreneurial clusters (Porter, 1998). Competition within a cluster necessitates the creation of a competitive advantage (Tsai et al., 2011), which is a quality signal of compliance with applicable privacy regulation. Gazel and Schwienbacher (2021) provide empirical evidence that location in a cluster reduces the risk of firm failure for FinTechs. We collected the data from the German company register.

Second, we consider the variable *legal capital*. This variable reflects the founder's dedication and readiness to make a notable investment in the own venture at an early stage of development (Hornuf et al., 2021b) and which can be interpreted as a quality signal of motivation and future success of business operations. In Germany, for the most common legal form of a limited liability company (the so-called *GmbH*), one needs to raise legal capital of at least 12,500 EUR at the time of incorporation. The dummy variable equals 1 if the minimum capital requirement of the underlying legal form amounts to more than 1 EUR and 0 otherwise. We derived this information from the German company register and imprints of the FinTech websites.

Third, we include number of *employees* as a proxy for FinTech companies' human capital and size (Hornuf et al., 2018a). *Employees* is a rank variable ranging from 1 to 5 and representing number of employees: 1-10, 11-50, 51-100, 101-1000 and above 1000. A larger number of employees usually means a more diversified team in terms of members' abilities and skills, resulting in venture success (Duchesneau and Gartner, 1990), which might also translate to compliance and legal aspects. For privacy-related aspects, Ramadorai et al. (2021) outline that larger firms tend to extract more data. Therefore, we proxy for firm size and human capital strength using the number of employees. We derived the data from BvD Dafne and complemented them with data from the Crunchbase database as well as

LinkedIn entries.

Fourth, we control for the age of the FinTech company during the particular data-collection period since its year of incorporation with the variable *firm age*. This variable serves as a proxy for a FinTech's stage of business (Hornuf et al., 2021b). We assume that established companies pay more attention to privacy aspects because they have more experience and available resources. Bakos et al. (2014) find for contracts in boilerplate language that consumers have more confidence in larger and older companies because they seem more credible and fair. We derive the year of incorporation from the German company register and respectively calculate it as the difference of the data collection period before and after the GDPR became binding.

We further include *industry dummies* to account for the diversity of business models. Our industry classification follows the FinTech taxonomy of Dorfleitner et al. (2017) with the segments and sub-segments (in parentheses): financing (donation-based crowdfunding, reward-based crowdfunding, crowdinvesting, crowdlending, credit and factoring), asset management (social trading, robo-advice, personal financial management, investment and banking), payments (alternative payment methods, blockchain and cryptocurrencies, other payment FinTechs) and other FinTechs (insurance, search engines and comparison websites, technology IT and infrastructure, other FinTechs). The categorization is based on FinTechs' business models in accordance with the functions and business processes of traditional banks. The business model provides first indications about the data processing of a specific FinTech because in a digitized industry, data are often at the core of the business model.

The variables *employees, legal capital, bankcooperation* and *city* are time-invariant. We collected all variables in this paper respectively before and after the GDPR became binding.

2.4.3 Methods

2.4.3.1 Textual analysis: preprocessing

We prepare the texts of the privacy statements using standard methods of text mining, including cleaning to remove white spaces, numbers, punctuation and other symbols. For the standardization analysis, we also need to consider that the language of the privacy statements is German. We therefore remove capitalization and apply stemming to the German language to reduce words to their root in order to consider different grammatical forms of the same word family. We delete stop words with the help of the German stop word list in the R package "lsa" (Wild, 2007) because stop words such as articles, conjunctions and frequently used prepositions do not convey additional meaning. Subsequently, we break the texts down into tokens that represent individual words and count their frequency within each text separately for both data-collection periods.

2.4.3.2 Readability

The GDPR refers to the comprehensibility of privacy statements in order to achieve transparency with "easy to understand, and [...] clear and plain language" (rec. 39 GDPR) and mentions "that it should be understood by an average member of the intended audience" (Article 29 Working Party, 2018). Readability is defined as the ease of understanding a text and is usually measured using formulas based on sentence length, syllables and word complexity. The most commonly used readability measures in academic literature are the Flesch reading ease score (Flesch, 1948) and the Gunning Fog Index (Gunning, 1952), both corresponding to the number of formal years of education required to comprehend a text. We investigate companies operating in Germany and because the privacy statements are often written in German, we address the variety of morphological and semantic richness by using metrics for or adapted to German.

First, we apply the Neue Wiener Sachtext formula by Bamberger and Vanecek (1984) using the formula

$$nWS = 0.1935 \cdot \frac{n_{wsy\geq 3}}{n_w} + 0.1672 \cdot ASL + 0.1297 \cdot \frac{n_{wchar\geq 6}}{n_w} - 0.0327 \cdot \frac{n_{wsy=1}}{n_w} - 0.875 \tag{2.1}$$

where $n_{wsy\geq 3}$ is the number of words with three syllables or more, ASL is the average sentence length (number of words / number of sentences), $n_{wchar\geq 6}$ is the number of words with 6 characters or more and $n_{wsy=1}$ is the number of words of one syllable.

Second, we calculate the simplified SMOG metric of McLaughlin (1969) adapted to the peculiarities of the German language as

$$SMOG\ German = \sqrt{Nw_{min3sy} \cdot \frac{30}{n_{st}} - 2} \tag{2.2}$$

where Nw_{min3sy} is the number of words with a minimum of three syllables and n_{st} is the number of sentences (Bamberger and Vanecek, 1984). While these formulas for determining readability are frequently used in the literature (Loughran and McDonald, 2016; Ramadorai et al., 2021), they are nevertheless often criticized (Loughran and McDonald, 2014). Regarding privacy statements, Singh et al. (2011) state that the measures take into account sentence complexity and word choice but no aspects that determine comprehension. To address these points of criticism, we additionally consider the variable *No. words*, defined as the logarithm of the total number of words in the privacy statements. We consider the variable as an alternative measure of the understandability and complexity of a text reflected in the time required to read the whole text.

2.4.3.3 Standardization

To test Hypothesis 2, we quantify the extent to which the texts of privacy statements are standardized by calculating common measures of text similarity and distance for dissimilarity. We apply the vector space model (VSM) of Salton et al. (1975) to convert texts into term frequency-vectors, which enables us to perform algebraic calculations. The accounting and finance literature often applies Cosine or Jaccard similarity to account for similarity (Cohen et al., 2020; Peterson et al., 2015).[8]

As a first similarity measure, we calculate the *Cosine similarity.* Because of the vector representation of the texts, we can calculate the cosine of the included angle. The *Cosine similarity* between two documents is defined as the scalar product of the two term-frequency vectors divided by the product of their Euclidean norms. The values range from 0 to 1 because term-frequency vectors of texts cannot be negative. A main property of the *Cosine similarity* is that it does not consider text length. A value close to 1 indicates the presence of pure boilerplate language. The second similarity measure we calculate is *Jaccard similarity*, defined as the quotient of the size of the intersection and the size of the corresponding union of two term-frequency representations. In contrast to *Cosine similarity*, for the *Jaccard similarity* each word occurs only once in the sample, and its frequency is not accounted for. For privacy statements, Ramadorai et al. (2021) use Cosine similarity to analyze industry-specific boilerplate, whereas Kaur et al. (2018) employ Jaccard similarity to measure keyword similarity. Besides the similarity measures, we calculate the two distance metrics *Euclidean distance* and *Manhattan distance. Euclidean distance* is the shortest distance between the two document vectors with the corresponding term weights. In contrast to *Euclidean distance*, *Manhattan distance* is the absolute distance between the two vectors. Unlike for the similarity measures, values of distance metrics close to 1 indicate no correspondence between the analyzed texts.

We calculate all the aforementioned similarity and distance measures pairwise for the privacy statement texts D_1 and D_2 of two different companies within one data-collection period. In the next step, to obtain one average similarity or distance-measure value for one company before and after the GDPR became binding, we calculate our similarity and distance measures in relation to the average privacy statement per period, analogous to the “centroid vector [or] the average policy" of Ramadorai et al. (2021), as

$$\overline{D} = \frac{\sum_{n=1}^{N} D_n}{N} \tag{2.3}$$

where $\overline{D}$ is the average value per year, $\sum D_n$ the sum of the similarity respective distance of one FinTech's document in relation to every other document, and N the number of

[8] For illustrative examples of Cosine and Jaccard similarity, see Cohen et al. (2020).

companies.

2.4.3.4 Empirical Approach

To test our Hypotheses 1, 2, 3a and 3b, we use a two-sided paired t-test to examine whether the mean values of readability, standardization, quantity of data processed, and transparency are significantly different for the periods before and after the GDPR became binding.

We test Hypotheses 4a, 4b and 5 in a multivariate setting. Because our dependent variables are fractional indices in the interval between 0 and 1, we estimate fractional probit regressions using quasi-maximum likelihood (Papke and Wooldridge, 1996).

We further explore in Hypothesis 4a and 4b determinants of the *data index* and *transparency index* in separate models before and after the GDPR became binding. To compare the obtained regression coefficients of non-linear models for the same sample of companies at two different points in time, we further conduct seemingly unrelated estimations (Zellner, 1962). Then, we perform Wald chi-square tests to test whether the coefficients differ across our analyzed periods. The validity of the tests is ensured by the previously performed estimation based on the stacking method with respect to the appropriate co-variance matrix of the estimators for the standard errors (Weesie, 1999) and was formerly successfully applied by Mac an Bhaird and Lucey (2010) and Laursen and Salter (2014).

2.5 Results

2.5.1 Sample

Figure 2.1 shows the graphical distribution of the companies in their sub-segments following the detailed FinTech taxonomy of Dorfleitner et al. (2017). Table 2.2 provides summary statistics for all our variables. Most of the companies in the sample operate in the crowdinvesting and alternative payments, insurance respective IT, technology and infrastructure sub-segments. Crowdinvesting can be a data-intensive sub-segment (Ahlers et al., 2015), whereas payment providers receive manifold payment data that can entail almost all possible information about a person. Moreover, insurance companies typically process health data, which are special categories of personal data (art. 9 GDPR). The descriptive statistics of *bankcooperation* indicate that, on average, 25.4% of FinTechs in our sample maintain a cooperation with a bank. The median of *No. investors* is 0, which indicates that less than half the companies in the sample have received external funds. The mean and median values of *employees*, around 2, indicate that most of our FinTechs are small companies employing 11 to 50 people. The variable *city* indicates that, on average, 48.6% of the analyzed FinTechs are located in a large city.

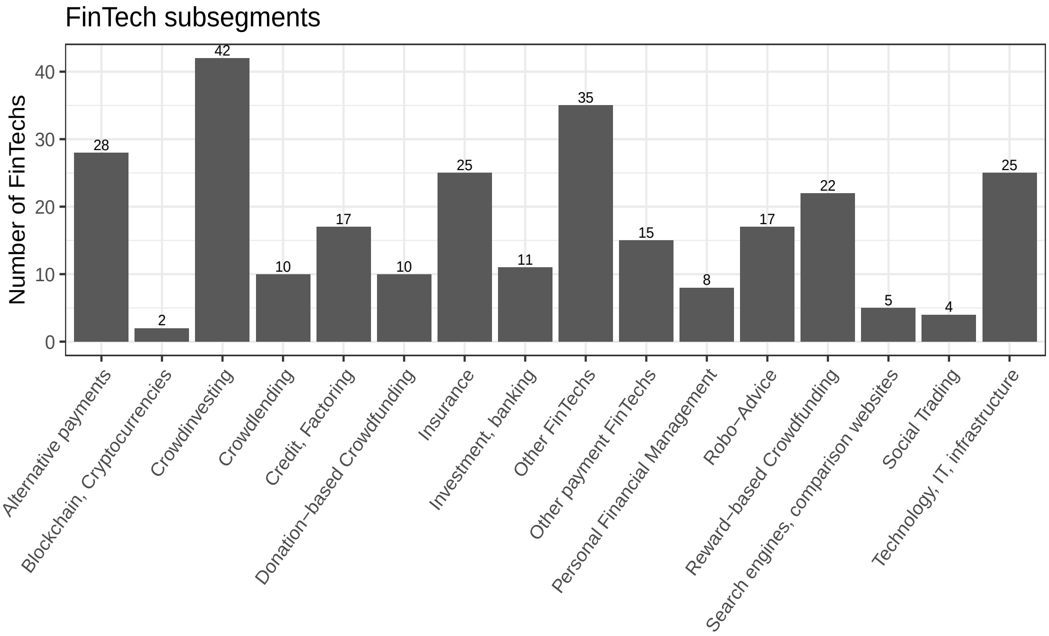


Figure 2.1: Frequency of occurrence of the FinTech sub-segments following the taxonomy of Dorfleitner et al. (2017), the bars represent the number of companies in each sub-segment. N=276.

2.5.1.1 Readability

The mean and median in combination with the quantiles of the readability metrics *Wiener Sachtext* and *SMOG German* increase slightly, which indicates that the readability of the privacy statements worsened after the GDPR became binding. In Table 2.3, two t-tests indicate a significant difference in means for both metrics (paired t-tests, $t = 2.569$ and $p < 0.05$, $t = 6.010$ and $p < 0.01$). The alternative readability proxy *No. words* shows a clear increase in any summary statistic, which indicates that the privacy statements contain more words and require more time to read. The increase is confirmed by a t-test on differences in means (paired t-test, $t = 15.017$, $p < 0.01$).

The cumulative distribution functions of all our variables considering readability are illustrated in Figure 2.2. A shift of the graph to the right indicates a worsening in readability from before (black) to after (grey) the GDPR became binding, which is evident for all our measures. These results are contrary to the GDPR's objective of clear and plain language. A discussion of the result for *Wiener Sachtext* and *SMOG German* requires a closer look at the method. Both metrics are mainly calculated based on word complexity and sentence length. In particular, word complexity is a critical issue for technical termini, which accompanies privacy-related legalese. Because the information content and quality regarding advanced technological topics can suffer from simpler language (Wachter, 2018).

Table 2.2: Descriptive statistics of all variables.

Variable	Mean	S.D.	Min	Q1	Median	Q3	Max
Legal capital	0.888	0.316	0.000	1.000	1.000	1.000	1.000
Bankcooperation	0.254	0.436	0.000	0.000	0.000	1.000	1.000
Employees	2.130	1.050	1.000	1.000	2.000	2.000	5.000
City	0.486	0.501	0.000	0.000	0.000	1.000	1.000
Firm age_pre	1.534	0.529	0.000	1.099	1.498	1.792	3.091
Firm age_post	1.749	0.436	0.693	1.386	1.701	1.946	3.136
No. investors_pre	0.714	1.047	0.000	0.000	0.000	1.400	4.000
No. investors_post	0.754	1.072	0.000	0.000	0.000	1.400	4.000
Wiener Sachtext_pre	13.654	0.915	10.113	12.988	13.739	14.332	17.270
Wiener Sachtext_post	13.860	1.127	9.888	13.149	13.866	14.625	17.225
SMOG German_pre	12.266	1.221	8.955	11.321	12.255	13.222	16.974
SMOG German_post	12.992	1.749	8.191	11.809	13.111	14.070	17.310
No. words_pre	7.102	0.864	2.890	6.796	7.252	7.601	8.970
No. words_post	7.866	0.867	2.944	7.453	7.959	8.443	9.622
Cosine similarity_pre	0.533	0.095	0.132	0.495	0.559	0.601	0.659
Cosine similarity_post	0.583	0.089	0.126	0.537	0.603	0.651	0.706
Jaccard similarity_pre	0.207	0.047	0.023	0.191	0.217	0.238	0.276
Jaccard similarity_post	0.227	0.044	0.014	0.214	0.240	0.255	0.280
Euclidean distance_pre	0.096	0.024	0.074	0.083	0.090	0.101	0.303
Euclidean distance_post	0.081	0.023	0.062	0.070	0.076	0.087	0.318
Manhattan distance_pre	1.312	0.136	1.132	1.226	1.275	1.350	1.901
Manhattan distance_post	1.255	0.125	1.097	1.178	1.229	1.286	1.934
Data Index_pre	0.206	0.103	0.000	0.125	0.200	0.275	0.575
Data Index_post	0.237	0.098	0.000	0.169	0.225	0.300	0.550
Transparency Index_pre	0.303	0.175	0.000	0.125	0.375	0.375	0.875
Transparency Index_post	0.295	0.158	0.000	0.125	0.250	0.375	0.875
Mimic Data Index	0.206	0.037	0.106	0.181	0.198	0.226	0.340
Mimic Transparency Index	0.303	0.065	0.175	0.254	0.290	0.335	0.425

Note: Descriptive statistics for all our variables, the abbrevation "_pre" indicates before and "_post" after the GDPR became binding. N=276. The variables are defined in Table 2.1.

Table 2.3: Paired two-sided t-test to test Hypotheses 1, 2, 3a, 3b.

	Pre-GDPR		*Post-GDPR*				
Variable	Mean	S.D.	Mean	S.D.	Diff.	t-stat	p-value
Wiener Sachtext	13.654	0.915	13.860	0.206	0.206	2.569	0.011*
SMOG German	12.266	1.221	12.992	1.749	0.726	6.010	0.000***
No. words	7.102	0.864	7.866	0.867	0.764	15.017	0.000***
Cosine similarity	0.533	0.095	0.583	0.090	0.050	8.606	0.000***
Jaccard similarity	0.207	0.047	0.227	0.044	0.020	6.880	0.000***
Euclidean distance	0.096	0.024	0.081	0.023	−0.015	−12.530	0.000***
Manhattan distance	1.312	0.136	1.255	0.125	−0.057	−7.074	0.000***
Data Index	0.206	0.103	0.237	0.098	0.031	5.940	0.000***
Transparency Index	0.303	0.175	0.295	0.158	−0.009	−0.904	0.367

Note: Paired two-sided t-test (significance level of 5%) to test Hypothesis 1 regarding readability, Hypothesis 2 regarding standardization and Hypotheses 3a and 3b regarding quantity of data processed and transparency. N=276. The variables are defined in Table 2.1.

It is not surprising that in the FinTech industry a more complex language has recently been used to describe the data processing of complex business models based on, for example, artificial intelligence or the blockchain technology. Our results for *No. words* are in line with Linden et al. (2020), who find in their before and after the GDPR comparison an increase in text length but no changes in sentence structure. Thus, our evidence supports Hypothesis 1.

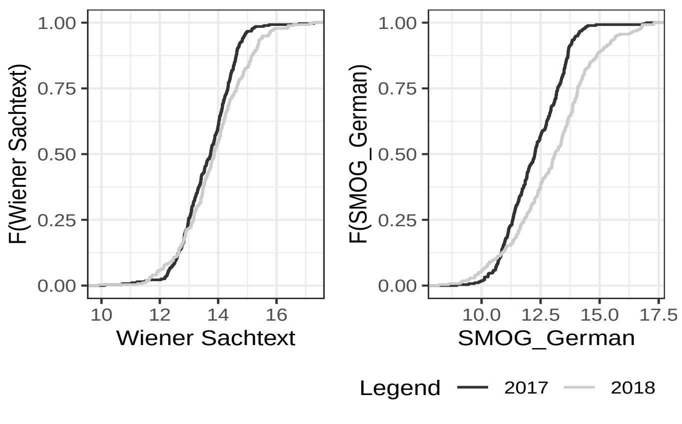


Figure 2.2: Cumulative distribution function for the readability measures *Wiener Sachtext, SMOG German* and *No. words* for before (2017, black) and after (2018, grey) the GDPR became binding. N=276. The variables are defined in Table 2.1.

2.5.1.2 Standardization

In this section, we test Hypothesis 2 on the increase of boilerplate language after the GDPR became binding. The similarity measures *Cosine similarity* and *Jaccard similarity* reveal a clear increase in mean and median. Both measures indicate an increase in boilerplate language, which is confirmed by a t-test for differences in means at conventional levels (paired t-tests, $t = 8.606$ and $p < 0.01$, $t = 6.880$ and $p < 0.01$). Consistent with the similarity metrics, we identify for the distance metrics *Euclidean distance* and *Manhattan distance* a decrease in means and medians, indicating an increase in the use of boilerplate language. The means are statistically significantly different before and after the GDPR became binding (paired t-tests, $t = -12.530$ and $p < 0.01$, $t = -7.074$ and $p < 0.01$). The standard deviation for all measures remains almost the same for both sample periods. Regarding all of our similarity and distance metrics, the first and third quantiles are far from the minima or maxima, illustrating that although some outliers exist, there is a tendency towards the mean and the median. In Figure 2.3, the cumulative distribution function of the similarity and distance measures illustrates a shift to more similar and therefore standardized language from before (grey) compared with after (black) the GDPR became binding.

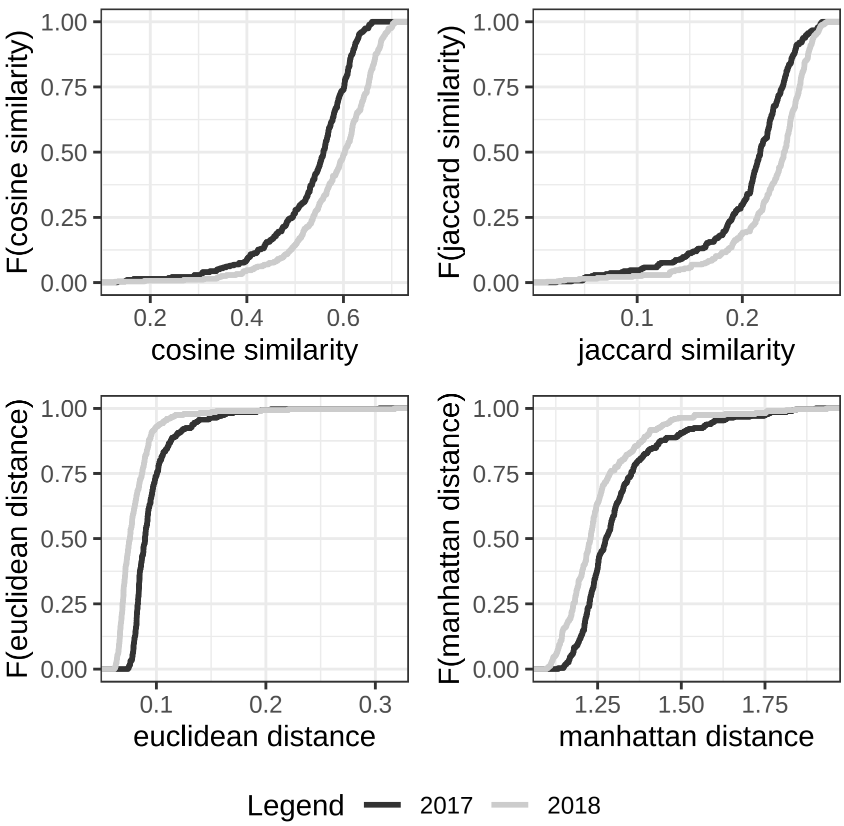


Figure 2.3: Cumulative distribution function for the similarity and distance measures *cosine similarity, jaccard similarity, euclidean distance* and *manhattan distance* for before (2017, black) and after (2018, grey) the GDPR became binding. N=276. The variables are defined in Table 2.1.

In sum, we find an increase in privacy statements' use of standardized language after the GDPR became binding. Companies appear to have chosen the path towards legal-safeguarding boilerplate policies to the detriment of their users. Overall, Hypothesis 2 receives support.

2.5.1.3 Quantity of data processed and transparency

In this section, we move from the analysis of the readability as the basic requirement for transparency to the actual transparency in terms of content of the privacy statements.[9] For the *data index*, we find an increase in the mean and median from before to after the GDPR became binding, which illustrates that companies state more often in their privacy statements post-GDPR that they process specific data. The difference is statistically significant in a t-test (paired t-test, $t = 5.940$, $p < 0.01$). Thus, we find supportive evidence for Hypothesis 3a. A closer look at all summary statistics emphasizes large divergences in the quantity of data processed between the individual companies. The *data index* minimum of 0 indicates that some firms do not state that they process any data. The range of the actual maximum value before and after the GDPR became binding indicates that even companies that process a lot of data are far from the maximum theoretical index value of 1.

For the *transparency index*, we find a small decrease in the mean and median. This finding suggests that, contrary to our Hypothesis 3b, companies' privacy practices have not improved in terms of transparency since the GDPR became binding. Note that there are companies in both periods reaching a maximum value of 0.875 for the *transparency index*, which indicates a high level of transparency. After performing a t-test on the mean, we find no statistically significant difference (paired t-test, $t = -0.940$, $p > 0.05$). Thus, we find no empirical support for Hypothesis 3b, that transparency has increased since the GDPR became binding. However, one must bear in mind that the FinTech industry operates in a highly competitive environment and is caught between the pressure to innovate and state-of-the-art data privacy. For this reason, it can be difficult for FinTechs to be fully transparent without losing their competitive edge to competitors. In contrast to our results, Linden et al. (2020) use different but closely related transparency measures and conclude that transparency has improved since the GDPR became binding but that privacy statements are far from fully transparent.

When considering the results of both indices, we conclude that since the GDPR became binding, FinTechs state that they process more data although they have not made efforts to enhance the transparency of privacy practices. Further, we identify large differences between individual companies. A possible explanation is that the FinTech industry as a whole is highly diverse and that the different business models require different intensities

[9] For detailed summary statistics of our disaggregated indices, we refer readers to Tables A.8 and A.9 in the Appendix.

of data processing. For example, crowdinvesting platforms process a lot of data. The projects and initiator data need to be assessed in detail before the funding. During the funding process, disclosure of more information about the project and the initiators has been identified as a success factor (Ahlers et al., 2015).

2.5.1.4 Determinants of the quantity of data processed and transparency

Table 2.4 shows the results for Hypotheses 4a and 4b on the effect of the number of investors and the existence of a bank cooperation on the quantity of data processed and transparency.

We find that before the GDPR became binding, the coefficient of *No. investors* is positive and significant at the 5% level for both indices, where a one-standard-deviation increase in *No. investors* is associated with a 55.9% increase in the *data index* in model (1) and a 41.2% increase in *transparency* in model (3). However, the effect and significance of the variable disappear for the period after the GDPR became binding in models (2) and (4). Before the GDPR became binding, the number of external investors had a positive effect on data-privacy compliance because it was positively related to the quantity of data processed and to transparency. Our results for *No. investors* provide partial support for Hypothesis 4a.

Further, none of our regression models yield a significant effect of *bankcooperation* on quantity of data processed or on transparency. Because of missing significances, we cannot provide further evidence for how external investors or cooperating banks influenced the implementation of the GDPR by FinTechs. Regarding *bankcooperation*, we find no empirical support for Hypothesis 4b.

The control variable *legal capital* has a significant positive influence on both indices for all models, which indicates that founders who invested more legal capital are also more dedicated to their business in terms of data privacy compliance. Wald tests for differences in coefficients before and after the GDPR became binding only show a significant difference for legal capital as a determinant of the *transparency index*. The coefficients for the *transparency index* are significantly different and lower after the GDPR (Wald chi-square test, $\chi^2 = 4.740$, $p < 0.05$). Thus, the effect of *legal capital* on transparency is stronger before the GDPR. This may be because before the GDPR became binding, only highly dedicated founders invested time in privacy compliance, whereas the GDPR made this issue the focus of every company. We consider variance inflation factors (VIF), reported in Table A.3 in the Appendix, and find no indications of multicollinearity for any of our model specifications.

Table 2.4: Seemingly unrelated fractional probit regressions to test Hypotheses 4a and 4b.

	Dependent variable:					
	Data Index			Transparency Index		
	Pre-GDPR	Post-GDPR	Wald-Test	Pre-GDPR	Post-GDPR	Wald-Test
	(1)	(2)	p-value	(3)	(4)	p-value
No. investors	0.055*	0.025	0.358	0.069*	0.042	0.568
	(0.024)	(0.022)		(0.034)	(0.031)	
Bankcooperation	0.031	−0.005	0.574	0.035	0.031	0.970
	(0.048)	(0.044)		(0.070)	(0.065)	
Legal capital	0.247***	0.107+	0.137	0.575***	0.268**	0.037*
	(0.071)	(0.061)		(0.095)	(0.092)	
City	−0.059	0.010	0.223	−0.045	0.075	0.137
	(0.042)	(0.038)		(0.058)	(0.056)	
Firm age	−0.025	−0.019	0.913	−0.061	−0.083	0.796
	(0.042)	(0.043)		(0.055)	(0.067)	
Employees	0.032	0.040+	0.814	0.006	0.033	0.604
	(0.024)	(0.022)		(0.035)	(0.037)	
Constant	−0.752***	−0.601***		−0.711***	−0.624***	
	(0.155)	(0.153)		(0.161)	(0.164)	
Industry Effects	Yes	Yes		Yes	Yes	
Observations	276	276	276	276	276	276
Log Likelihood	−97.239	−103.970		−115.858	−115.384	

Note: Seemingly unrelated fractional probit estimations to test Hypotheses 4a and 4b regarding determinants on the quantity of data processed and transparency, Wald-Test (significance level of 5%) with p-values to compare equality of coefficients of models (1)(2) and (3)(4), numbers in parentheses are robust standard errors. The variables are defined in Table 2.1. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

2.5.1.5 Mimicking behavior

Table 2.5 reports the results for Hypothesis 5, which considers mimicking behavior regarding data privacy compliance among industry peers.

Table 2.5: Fractional probit regression to test Hypothesis 5.

	Dependent variable:	
	Data Index Post-GDPR	Transparency Index Post-GDPR
	(1)	(2)
No. investors	0.018	0.024
	(0.023)	(0.033)
Bankcooperation	0.026	0.053
	(0.045)	(0.061)
Legal capital	0.110+	0.174+
	(0.064)	(0.102)
City	0.021	0.079
	(0.039)	(0.054)
Firm age	−0.030	−0.044
	(0.034)	(0.052)
Employees	0.027	0.026
	(0.022)	(0.037)
Mimic Data Index	1.613**	
	(0.554)	
Mimic Transparency Index		2.027***
		(0.427)
Constant	−1.189***	−1.373***
	(0.145)	(0.176)
Industry Effects	No	No
Observations	276	276
Log pseudolikelihood	−150.6126	−165.5007

Note: Fractional probit regression to test Hypothesis 5 regarding mimicking behavior, numbers in parentheses are robust standard errors. The variables are defined in Table 2.1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

In Table 2.5 model (1), we find a positive significant effect of the *mimic data index* on the 1% significance level, in which a one-standard-deviation increase in *mimic data index* is associated with a 58.6% increase in the *data index* relative to the average. In model (2), we find a highly significant impact of the *mimic transparency index* on the *transparency index*, in which a one-standard-deviation increase in the explanatory variables leads to a 75.27% increase in the dependent variable relative to the average. The results indicate a strong mimicking behaviour among industry peers in terms of data privacy compliance, because a higher industry average for both indices before the GDPR became binding accompanies more data processed and greater transparency for a specific company.[10] Thus, the conjecture that FinTechs mimic the privacy statements of their industry peers is supported by our evidence. As for our control variables, we find a weak statistically positive effect for *legal capital* for both indices. In sum, we find supportive evidence for Hypothesis 5 on mimicking behavior.

2.6 Robustness

Finally, we perform robustness checks and estimate alternative specifications to test the validity of our results.

2.6.1 Sub-sample: exclusion of mature FinTechs

To test for the influence of more mature FinTechs, we exclude companies, like Hornuf et al. (2021a), that employ more than 1000 people or that were founded at least 10 years before our first data-collection period. More experienced and larger companies have more free resources to address legal issues. Especially regarding boilerplate and mimicking behavior, it could be argued that larger or older firms are role models for their immediate industry peers and whose privacy practices are mimicked. When excluding these FinTechs, 249 companies remain in the sample. In Table A.4 in the Appendix, we report summary statistics for the text-feature analysis and find patterns remarkably similar to those for the whole sample analyzed in the "Results" section. For the regression estimates in Tables A.5 and A.6 in the Appendix, we observe no changes in signs and only small changes in significance of the coefficients. Therefore, we note that it is unlikely that more mature FinTechs drive our results.

2.6.2 Pooled OLS with GDPR interaction

To verify our results for the year-wise estimations and post-estimation tests in the seemingly unrelated estimations in the "Results" section, we run an OLS regression with the interaction dummy variable *GDPR*. We estimate the OLS regression to simplify the regression model for the link function in the prior probit specification and pool our observations in a single model with the GDPR interaction to evaluate the effect of the policy intervention simultaneously. The results are reported in Table A.7 in the Appendix and mostly show

[10] In unreported analysis, we estimate the same model using a mimicking variable based on segment-level averages of finance, asset management, payments and other FinTechs. Interestingly, we find for that specification no statistically significant coefficients and thus conclude that the less detailed categorization fails to depict commonalities in business models, data processing and consequently mimicking behavior.

similar patterns in terms of signs and significance of the coefficients compared with the prior model specifications. Additionally, we find that the dummy variable *GDPR* itself has a positive significant influence on level of transparency.

2.6.3 Causality

Endogeneity problems in empirical studies can come in a variety of forms. Reverse causality and simultaneity are among the most relevant. In this study, the results in Tables 2.4 and 2.5 could, for example, be affected by reverse causality. However, when considering the significant variable *legal capital*, it can be argued that the decision for the *legal capital* is made at the moment the company is founded, while the decision for the dependent variable, the *data and transparency index*, is made at a later stage when the company begins operations. As for simultaneity, variables that are potentially missing should, for example, correlate with *firm age*, which is not significant though. This gives us some confidence that endogeneity is not an obvious problem in our analysis.

2.7 Conclusion

The theoretical framework of this study is embedded in the general legal principle for data processing, namely transparency (art. 5(1)a GDPR). We empirically study the degree of implementation of the GDPR by FinTech companies operating in Germany. For this purpose, we analyze the privacy statements of 276 FinTechs before and after the GDPR became binding. We use methods from text analysis, extend our findings using a content-based approach, and link this to FinTech company- and industry-specific determinants.

With regard to the text-feature analysis, we document a decrease in readability in conjunction with substantially longer texts and more time required to read the privacy statements. The FinTechs appear to safeguard themselves with exact technical and legal termini and comprehensive statements instead of the user comprehension required by the GDPR. We further find indications of an increase in standardized legal language built on the literature of boilerplate after the GDPR became binding, reducing the informational content that users can draw from the texts. These findings contradict the basic requirements for transparency of the GDPR. Further, we analyze the quantity of data processed, the actual transparency of privacy statements, and their determinants. We document a significant increase in the quantity of data processed but find no significant changes in the level of transparency. The number of external investors positively influences the quantity of data processed and transparency solely before the GDPR became binding. Regarding cooperation with a bank, we find no significant effects in any specification. Legal capital that we interpret as *ex-ante* founder team dedication is positively related to the level of privacy and is particularly relevant for transparency before the GDPR became binding. These results underline that before the GDPR became binding, externally induced pressure of investors and internal engagement of the founders resulted in better privacy practices. However, the results vanish after the GDPR became binding, as the GDPR made all FinTechs act to ensure data privacy.

We ask whether it is possible for a user to give informed consent (art. 7 GDPR) if they cannot transparently capture the language respective to the content of privacy statements. This raises the question of whether FinTech companies have implemented the essential provisions of the GDPR and whether the regulation has achieved its goal. The answer is broadly no. Looking at the question from a company perspective, however, one has to consider whether a company can ever be fully GDPR compliant without seriously restricting its business activities. This is particularly relevant for a data-intensive and competitive industry like the FinTech industry. From the perspective of regulators, one might ask whether the GDPR is deficient in the sense that the financial industry needs to simplify the language of privacy statements so that laypeople can understand what information is being processed and how. We do not assume that laypersons will actually read privacy statements and enforce their rights (Strahilevitz and Kugler, 2016), which would be associated with far too high transaction costs. Rather, as with securities prospectuses, professional market participants such as data protection authorities are usually the addressees of privacy statements. They have the task of preparing the information and communicating it to the broader audience of customers (Firtel, 1999). So far, however, no comprehensive measures are known in which European or national data protection authorities have carried out extensive benchmarking of privacy statements. Tools supported by artificial intelligence in particular could help here, enabling consumers to have privacy statements checked online. They could examine the privacy statements for content and summarize them in simplified language.

We also provide evidence that mimicking behavior in terms of FinTech industry pressure positively influences data-privacy compliance after the GDPR became binding, which indicates that the GDPR gave companies an incentive to adopt their direct industry peers' data-processing or privacy statements. This raises the question of whether FinTech companies can gain a competitive advantage over their peers by improving their privacy policies. The current literature is inconclusive about whether high quality and easy to read privacy statements lead to a competitive advantage. Even if privacy statements are read by the customers, the one-sidedness of privacy statements will, however, most likely not trigger a race to the top (Marotta-Wurgler, 2008; Marotta-Wurgler and Chen, 2012). This would perhaps only be the case if professional market participants make privacy statements easily comparable and accessible to a broad public. Even in such a scenario, an inferior standard could also prevail if network effects support the demand for a common, potentially inferior standard agreement (Engert and Hornuf, 2018).

Despite FinTech companies' imperfect implementation of the GDPR, our results nevertheless point to managerial recommendations. Our analysis of mimicking behavior shows, among other things, that companies take heed of the data privacy behavior of others. If data protection authorities and the media make the quality of privacy statements indeed transparent and easily accessible in the future, this could eventually lead to competition and a race to the top in privacy statement content. To excel in this competition, companies not only need to be compliant with the GDPR, but may also need to innovate in how privacy statements are agreed on. For example, users could actively give up parts of their data privacy in exchange for better prices or more usage rights, and conversely pay more to maintain greater data privacy. As is well known from the literature (Hillebrand et al., 2023), more transparency also leads to more trust and reputation gains for companies. For example, easy-to-click menus could help users prevent companies from sharing personal information with certain other companies when it is not strictly necessary for the performance of a contract. Here technical possibilities could help to enable FinTechs and consumers with a corresponding implementation.[11] Finally, the processing and forwarding of data could also be prepared and standardized in tabular form. However, standardization would require coordination among the companies in the FinTech industry and possibly new legislative initiatives.

Our article has limitations. We mainly refer to the privacy practices that companies declare in their privacy statements, and thus to the supply side of privacy (Ramadorai et al., 2021). Consumers must accept the terms for data processing if they want to use a service or product (Aridor et al., 2020). One avenue for further research is to compare what companies state in their privacy statements with the privacy practices they actually pursue. The results regarding transparency rely on our variable construction. Other approaches and methods can therefore yield different outcomes and insights. Similarly to Goldberg et al. (2021), we can only provide early evidence relating to our data-collection period shortly after the GDPR became binding in May 2018 and how the analyzed companies implemented the regulation at this point.

Finally, our article has practical implications. Legislators as well as policymakers in the EU and other countries that have adopted a privacy regulation related to the GDPR can now see the implications and the unintentional consequences of the regulation. This may pave the way towards future readjustment of the GDPR or give more practical guidance on

[11]There are already numerous tools that help companies to implement data privacy. See, for example, https://www.iitr.de/index.php, https://www.circle-unlimited.com/solutions/contracts/data-protection-management, https://compliance-aspekte.de/en/solutions/dsms/, https://www.dsgvo.tools, https://www.datenschutzexperte.de/dsgvo-tool/, and https://trusted.de/dsgvo-software. The providers of these tools could also extend them in such a way that a negotiation process about data transfer between companies and customers is facilitated.

how to create privacy statements to ensure compliance with the applicable legal standards. Further, our study emphasizes the importance of companies making greater efforts to implement effective privacy practices and communicate them to users in order to benefit from the opportunity to build a competitive advantage.

Appendix

Table A.1: Correlation matrix pre-GDPR.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Data Index	1							
Transparency Index	0.528	1						
No. investors	0.199	0.086	1					
Legal capital	0.185	0.256	0.065	1				
City	−0.016	−0.017	0.320	0.047	1			
Firm age	−0.002	−0.043	0.106	0.115	−0.005	1		
Bankcooperation	0.071	0.024	0.309	0.102	0.084	0.092	1	
Employees	0.205	0.078	0.577	0.066	0.183	0.235	0.198	1

Note: Correlation matrix for the data collection period before the GDPR became binding. The included variables correspond to the regression estimations in Table 2.4. N=276. The variables are defined in Table 2.1.

Table A.2: Correlation matrix post-GDPR.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Data Index	1									
Transparency Index	0.501	1								
No. investors	0.150	0.097	1							
Legal capital	0.114	0.110	0.061	1						
City	0.073	0.097	0.316	0.047	1					
Firm age	−0.020	−0.047	0.079	0.116	−0.008	1				
Bankcooperation	0.072	0.052	0.308	0.102	0.084	0.083	1			
Employees	0.156	0.104	0.576	0.066	0.183	0.244	0.198	1		
Mimic Data Index	0.212	0.151	0.048	0.001	−0.022	−0.103	−0.070	0.105	1	
Mimic Transparency Index	0.141	0.262	−0.129	−0.044	−0.091	−0.137	−0.135	−0.027	0.765	1

Note: Correlation matrix for the data collection period after the GDPR became binding. The included variables correspond to the regression estimations in Table 2.4 and Table 2.5. N=276. The variables are defined in Table 2.1.

Table A.3: Variance inflation factors.

	VIF1	VIF2	VIF3	VIF4	VIF5	VIF6
No. investors	1.85	1.85	1.85	1.88	1.71	1.73
Legal capital	1.19	1.19	1.19	1.19	1.02	1.02
City	1.18	1.18	1.18	1.18	1.12	1.12
Firm age	1.26	1.26	1.26	1.28	1.09	1.10
Bankcooperation	1.28	1.28	1.28	1.28	1.13	1.13
Employees	1.76	1.76	1.76	1.80	1.60	1.58
Mimic Data Index					1.04	
Mimic Transparency Index						1.06

Note: Variance inflation factors, VIF1-VIF4 correspond to Table 2.4 and models (1)-(4), VIF5-VIF6 correspond to Table 2.5 and models (1) and (2). The reported VIFs provide no indications for multicollinearity. N=276. The variables are defined in Table 2.1.

Table A.4: Descriptive statistics and paired t-test without mature FinTechs.

	Pre-GDPR			*Post-GDPR*					
Variable	Mean	SD	Median	Mean	SD	Median	Diff.	t-stat.	p-value
Wiener Sachtext	13.655	0.894	13.730	13.847	1.152	13.855	0.192	2.254	0.025*
SMOG German	12.283	1.221	12.247	12.973	1.770	13.0923	0.690	5.395	0.000***
No. words	7.085	0.848	7.228	7.856	0.875	7.975	0.771	14.414	0.000***
Cosine similarity	0.538	0.090	0.562	0.585	0.089	0.604	0.047	7.979	0.000***
Jaccard similarity	0.209	0.046	0.217	0.228	0.044	0.241	0.019	6.325	0.000***
Euclidean distance	0.096	0.024	0.089	0.081	0.024	0.076	−0.015	−11.744	0.000***
Manhattan distance	1.306	0.132	1.271	1.252	0.126	1.228	−0.054	−6.584	0.000***

Note: Sub-sample analysis, excluding mature FinTechs, summary statistics and paired two-sided t-tests (significance level of 5%) regarding the text-based variables. N=249. The variables are defined in Table 2.1.

Table A.5: Seemingly unrelated fractional probit regression without mature FinTechs.

	Dependent variable:					
	Data Index			Transparency Index		
	Pre-GDPR	Post-GDPR	Wald-Test	Pre-GDPR	Post-GDPR	Wald-Test
	(1)	(2)	p-value	(3)	(4)	p-value
No. investors	0.036	0.021	0.671	0.076*	0.037	0.460
	(0.025)	(0.025)		(0.035)	(0.036)	
Bankcooperation	0.011	−0.018	0.684	0.038	0.028	0.917
	(0.051)	(0.047)		(0.073)	(0.070)	
Legal capital	0.280***	0.130+	0.144	0.637***	0.286**	0.030
	(0.077)	(0.067)		(0.120)	(0.108)	
City	−0.048	0.007	0.373	−0.029	0.073	0.243
	(0.045)	(0.042)		(0.060)	(0.059)	
Firm Age	−0.012	−0.001	0.880	−0.101	−0.099	0.988
	(0.055)	(0.056)		(0.070)	(0.078)	
Employees	0.028	0.040	0.740	−0.016	0.039	0.355
	(0.026)	(0.026)		(0.042)	(0.042)	
Constant	−0.827***	−0.725***		−0.594**	−0.685***	
	(0.173)	(0.172)		(0.182)	(0.180)	
Industry Effects	Yes	Yes		Yes	Yes	
Observations	249	249	249	249	249	249
Log Likelihood	−87.899	−94.058		−104.089	−104.242	

Note: Sub-sample analysis, excluding mature FinTechs, seemingly unrelated fractional probit regression regarding determinants of the quantity of data processed and transparency, Wald-Test (significance level of 5%) p-values to compare equality of coefficients of models (1)(2) and (3)(4), numbers in parentheses are robust standard errors. The variables are defined in Table 2.1. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$.

Table A.6: Fractional probit regression without mature FinTechs.

	Dependent variable:	
	Data Index Post-GDPR	Transparency Index Post-GDPR
	(1)	(2)
No. investors	0.008	0.017
	(0.026)	(0.038)
Bankcooperation	0.010	0.043
	(0.048)	(0.065)
Legal capital	0.132+	0.179
	(0.070)	(0.115)
City	0.019	0.069
	(0.042)	(0.058)
Firm age	−0.007	−0.062
	(0.045)	(0.061)
Employees	0.031	0.036
	(0.026)	(0.044)
Mimic Data Index	1.402*	
	(0.715)	
Mimic Transparency Index		1.904***
		(0.440)
Constant	−1.190***	−1.322***
	(0.179)	(0.192)
Industry Effects	No	No
Observations	249	249
Log pseudolikelihood	−136.1568	−149.3130

Note: Sub-sample analysis, excluding mature FinTechs, fractional probit regression regarding mimicking behavior, numbers in parentheses are robust standard errors. The variables are defined in Table 2.1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A.7: Pooled OLS regression with GDPR interaction.

	Dependent variable:	
	Data Index	Transparency Index
	(1)	(2)
GDPR	0.052	0.046
	(0.036)	(0.061)
No. investors	0.016*	0.023+
	(0.007)	(0.012)
GDPR x No. investors	−0.008	−0.009
	(0.010)	(0.016)
Bankcooperation	0.001	0.002
	(0.014)	(0.023)
GDPR x Bankcooperation	0.006	0.016
	(0.019)	(0.031)
Legal capital	0.060***	0.175***
	(0.017)	(0.030)
GDPR x Legal capital	−0.025	−0.092*
	(0.025)	(0.044)
City	−0.019	−0.015
	(0.012)	(0.020)
GDPR x city	0.024	0.043
	(0.017)	(0.027)
Firm age	−0.005	−0.023
	(0.012)	(0.018)
GDPR x Firm age	0.001	0.003
	(0.017)	(0.027)
Employees	0.012+	0.003
	(0.007)	(0.012)
GDPR x Employees	−0.003	0.004
	(0.010)	(0.017)
Constant	0.240***	0.248***
	(0.041)	(0.046)
Industry Effects	Yes	Yes
Observations	552	552
R^2	0.187	0.199
Adj. R^2	0.143	0.156

Note: Pooled OLS regression with GDPR interaction, including the dummyvariable *GDPR* to take into account the effects of the GDPR, numbers in parentheses are robust standard errors. The variables are defined in Table 2.1. $^{+}$p<0.1; *p<0.05; **p<0.01; ***p<0.001.

Table A.8: Composition and descriptive statistics of *Data Index* and *Transparency Index* pre-GDPR.

Variable	Mean	SD	Min	Q1	Median	Q3	Max
Data index							
Name	0.678	0.468	0.000	0.000	1.000	1.000	1.000
Gender	0.116	0.321	0.000	0.000	0.000	0.000	1.000
Title	0.036	0.187	0.000	0.000	0.000	0.000	1.000
Language	0.011	0.104	0.000	0.000	0.000	0.000	1.000
Identifier	0.098	0.298	0.000	0.000	0.000	0.000	1.000
Password	0.145	0.353	0.000	0.000	0.000	0.000	1.000
Age	0.326	0.470	0.000	0.000	0.000	1.000	1.000
Place of birth	0.080	0.271	0.000	0.000	0.000	0.000	1.000
Address	0.572	0.496	0.000	0.000	1.000	1.000	1.000
E-mail address	0.612	0.488	0.000	0.000	1.000	1.000	1.000
Phone number	0.322	0.468	0.000	0.000	0.000	1.000	1.000
Residence city	0.029	0.168	0.000	0.000	0.000	0.000	1.000
Residence country	0.051	0.220	0.000	0.000	0.000	0.000	1.000
Marital status	0.040	0.196	0.000	0.000	0.000	0.000	1.000
Occupation	0.054	0.227	0.000	0.000	0.000	0.000	1.000
Bank	0.250	0.434	0.000	0.000	0.000	0.200	1.000
PIN	0.011	0.104	0.000	0.000	0.000	0.000	1.000
Income	0.040	0.196	0.000	0.000	0.000	0.000	1.000
Tax residency	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Social security number	0.011	0.104	0.000	0.000	0.000	0.000	1.000
Tax ident number	0.040	0.196	0.000	0.000	0.000	0.000	1.000
Driving license	0.007	0.085	0.000	0.000	0.000	0.000	1.000
Passport, registration	0.069	0.254	0.000	0.000	0.000	0.000	1.000
Graduation, qualification	0.011	0.104	0.000	0.000	0.000	0.000	1.000
Insurance	0.033	0.178	0.000	0.000	0.000	0.000	1.000
IP-address	0.141	0.349	0.000	0.000	0.000	0.000	1.000
GPS, location	0.029	0.168	0.000	0.000	0.000	0.000	1.000
Personal data published	0.149	0.356	0.000	0.000	0.000	0.000	1.000
Personal data transfer	0.158	0.365	0.000	0.000	0.000	0.000	1.000
Social Plugins, third party	0.525	0.500	0.000	0.000	1.000	1.000	1.000
Behavior, usage, movement	0.967	0.178	0.000	1.000	1.000	1.000	1.000
Google Analytics	0.826	0.380	0.000	1.000	1.000	1.000	1.000
Health	0.014	0.120	0.000	0.000	0.000	0.000	1.000
Religion	0.004	0.060	0.000	0.000	0.000	0.000	1.000
Nationality	0.083	0.277	0.000	0.000	0.000	0.000	1.000
Picture	0.072	0.260	0.000	0.000	0.000	0.000	1.000
Conversation record	0.004	0.060	0.000	0.000	0.000	0.000	1.000
Signature	0.014	0.120	0.000	0.000	0.000	0.000	1.000
Data Index	0.206	0.103	0.000	0.125	0.200	0.275	0.575
Transparency index							
Data	0.395	0.490	0.000	0.000	0.000	1.000	1.000
Purpose	0.859	0.349	0.000	1.000	1.000	1.000	1.000
Storage	0.489	0.501	0.000	0.000	0.000	1.000	1.000
Avoid	0.033	0.178	0.000	0.000	0.000	0.000	1.000
Opt-in	0.029	0.168	0.000	0.000	0.000	0.000	1.000
Pseudo	0.014	0.120	0.000	0.000	0.000	0.000	1.000
Third	0.113	0.317	0.000	0.000	0.000	0.000	1.000
Third data	0.498	0.501	0.000	0.000	0.000	1.000	1.000
Transparency Index	0.303	0.175	0.000	0.125	0.375	0.375	0.875

Note: Composition and descriptive statistics of *Data Index* and *Transparency Index* before the GDPR became binding. N=276. The variables are defined in Table 2.1.

Table A.9: Composition and descriptive statistics of *Data Index* and *Transparency Index* post-GDPR.

Statistic	Mean	SD	Min	Q1	Median	Q3	Max
Data index							
Name	0.768	0.423	0.000	1.000	1.000	1.000	1.000
Gender	0.192	0.395	0.000	0.000	0.000	0.000	1.000
Title	0.054	0.227	0.000	0.000	0.000	0.000	1.000
Language	0.014	0.120	0.000	0.000	0.000	0.000	1.000
Identifier	0.105	0.307	0.000	0.000	0.000	0.000	1.000
Password	0.199	0.400	0.000	0.000	0.000	0.000	1.000
Age	0.330	0.471	0.000	0.000	0.000	1.000	1.000
Place of birth	0.123	0.329	0.000	0.000	0.000	0.000	1.000
Address	0.580	0.495	0.000	0.000	1.000	1.000	1.000
E-mail address	0.790	0.408	0.000	1.000	1.000	1.000	1.000
Phone number	0.486	0.501	0.000	0.000	0.000	1.000	1.000
Residence city	0.025	0.158	0.000	0.000	0.000	0.000	1.000
Residence country	0.040	0.196	0.000	0.000	0.000	0.000	1.000
Marital status	0.043	0.204	0.000	0.000	0.000	0.000	1.000
Occupation	0.065	0.247	0.000	0.000	0.000	0.000	1.000
Bank	0.301	0.459	0.000	0.000	0.000	1.000	1.000
PIN	0.011	0.104	0.000	0.000	0.000	0.000	1.000
Income	0.033	0.178	0.000	0.000	0.000	0.000	1.000
Tax residency	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Social security number	0.004	0.060	0.000	0.000	0.000	0.000	1.000
Tax ident number	0.054	0.227	0.000	0.000	0.000	0.000	1.000
Driving license	0.007	0.085	0.000	0.000	0.000	0.000	1.000
Passport, registration	0.116	0.321	0.000	0.000	0.000	0.000	1.000
Graduation, qualification	0.007	0.085	0.000	0.000	0.000	0.000	1.000
Insurance	0.018	0.134	0.000	0.000	0.000	0.000	1.000
IP-address	0.366	0.483	0.000	0.000	0.000	1.000	1.000
GPS, location	0.025	0.158	0.000	0.000	0.000	0.000	1.000
Personal data published	0.185	0.389	0.000	0.000	0.000	0.000	1.000
Personal data transfer	0.163	0.370	0.000	0.000	0.000	0.000	1.000
Social Plugins, third party	0.638	0.482	0.000	0.000	1.000	1.000	1.000
Behavior, usage, movement	0.949	0.220	0.000	1.000	1.000	1.000	1.000
Google Analytics	0.808	0.395	0.000	1.000	1.000	1.000	1.000
Health	0.014	0.120	0.000	0.000	0.000	0.000	1.000
Religion	0.007	0.085	0.000	0.000	0.000	0.000	1.000
Nationality	0.101	0.302	0.000	0.000	0.000	0.000	1.000
Picture	0.087	0.282	0.000	0.000	0.000	0.000	1.000
Conversation record	0.011	0.104	0.000	0.000	0.000	0.000	1.000
Signature	0.007	0.085	0.000	0.000	0.000	0.000	1.000
Data Index	0.237	0.098	0.000	0.169	0.225	0.300	0.550
Transparency index							
Data	0.279	0.449	0.000	0.000	0.000	1.000	1.000
Purpose	0.920	0.271	0.000	1.000	1.000	1.000	1.000
Storage	0.406	0.492	0.000	0.000	0.000	1.000	1.000
Avoid	0.007	0.085	0.000	0.000	0.000	0.000	1.000
Opt-in	0.051	0.220	0.000	0.000	0.000	0.000	1.000
Pseudo	0.051	0.220	0.000	0.000	0.000	0.000	1.000
Third	0.076	0.266	0.000	0.000	0.000	0.000	1.000
Third data	0.569	0.496	0.000	0.000	1.000	1.000	1.000
Transparency Index	0.295	0.158	0.000	0.125	0.250	0.375	0.875

Note: Composition and descriptive statistics of *Data Index* and *Transparency Index* after the GDPR became binding. N=276. The variables are defined in Table 2.1.

Acknowledgements

We thank Fabio Bertoni, Engin Iyidogan, Youngjin Yoo, and the participants of the 7th Crowdinvesting Symposium (HU Berlin), the 5th European Alternative Finance Research Conference (University of Utrecht), the 3rd Machine Lawyering Conference (Chinese University of Hong Kong), the European Privacy InfoSec & Compliance Summit (EPIC) Summit (Munich), and the SKEMA-ESSEC 2022 Finance Workshop: FinTech and Decentralized Finance (SKEMA Business School, Nice) for their helpful comments and suggestions. Gregor Dorfleitner gratefully acknowledges the financial support by the Deutsche Bundesbank.

Chapter 3

Signaling in the Market for Security Tokens

This research project is joint work with Ralf Laschinger (University of Regensburg).

Abstract Security token offerings (STOs) are a new means for ventures to raise funding, where digital tokens are issued as regulated investment products on the blockchain. We study STO market outcomes in the primary and secondary markets for security tokens and examine the associated determinants in the context of signaling theory. We analyze success determinants of 138 STOs and find that a pre-sale and the announcement of token transferability are positively related to the funding success and serve as positive quality signals for investors to overcome information asymmetries. We examine 108 security tokens traded on centralized and decentralized exchanges related to the rapidly evolving area of decentralized finance. There is hardly any underpricing in the market, and it is positively associated with the crypto market sentiment as an external signal. When traded on the secondary market, security tokens generate both extremely positive and negative returns for various short-term time horizons. We disentangle the liquidity situation in the market between centralized and decentralized exchanges and find that decentralized marketplaces are less liquid and offer lower barriers to entry, which is an indication of slow market completion.

Keywords Security Token Offering, Blockchain, Signaling, STO, Decentralized Finance

JEL G24, K22, L26, M13, O31

3.1 Introduction

Advances in digitization and information technology have changed and transformed the financial industry fundamentally. Traditional financial institutions and banks are losing their supremacy as new market entrants and emerging technologies supersede or replace their role as financial intermediaries. Distributed ledger technology (DLT) and the blockchain, as its most common sub-type, enable the digitization of any asset class as tokens and are paving the way toward future financial markets. Digital security tokens are issued through token offerings on the blockchain, which represent an innovative funding mechanism in entrepreneurial finance. Once the token offering has taken place, the tokens can be traded on the secondary market.

In this study, we examine how signaling affects the behavior of market participants in both the pre-and post-STO phases to provide a holistic picture of the entire market. In particular, we study STO market outcomes such as STO funding success, underpricing, returns, liquidity, and various internal and external signals as determinants. Since the mechanisms and issuance processes are completely different because of blockchain technology, it is worth investigating whether signaling and related theories known from traditional capital markets also apply to the security token market.

The first tokens issued in the year 2013 were utility tokens sold through an initial coin offering (ICO). Utility tokens entail consumption rights for products or services. After a boom period in 2017 and 2018, the initial popularity of ICOs declined because of the lack of investor protection and many fraudulent activities, causing a negative market sentiment (Momtaz et al., 2019). As a result, security tokens issued through security token offerings (STOs) have since emerged as innovative investment products (Lambert et al., 2022). Security tokens represent shares of ownership in corporate equity, fixed income, investment funds, commodities, or less liquid asset classes such as real estate or fine art. Due to the classification as conventional securities and the resulting regulatory requirements, they are considered the regulatory-compliant successors to utility tokens. This new form of venture financing has several advantages: companies can easily reach a large investor base while reducing transaction costs. Moreover, clearing and settlement take place quickly and at any time, transparency regarding the transactions is achieved through the blockchain, and fractionalization enables investments in less liquid asset classes with high entry barriers (Ante and Fiedler, 2020; Lambert et al., 2022). The interoperability of the blockchain solves the previous problem of lack of compatibility between different systems or databases and enables self-custody of any tokenized asset on one platform (Momtaz, 2023). Another major advantage of STOs is the potential liquidity provided through the possibility to transfer and trade tokens on secondary markets. As a result, security tokens combine the benefits of the underlying technology with the legal protection of conventional securities.

Prior studies on ICOs analyzed success determinants (Adhami et al., 2018; Amsden and Schweizer, 2019; Fisch, 2019; Howell et al., 2020; Roosenboom et al., 2020), investor characteristics and motives (Boreiko and Risteski, 2021; Fahlenbrach and Frattaroli, 2021; Fisch et al., 2021; Hackober and Bock, 2021) or the informative disclosure and language of white papers (Florysiak and Schandlbauer, 2022; Thewissen et al., 2022). Other studies emphasize the post-ICO performance of tokens, such as underpricing (Chanson et al., 2018; Felix and von Eije, 2019) and/or short-term returns (Benedetti and Kostovetsky, 2021; Fisch and Momtaz, 2020; Lyandres et al., 2022; Momtaz, 2021a). However, due to the security and regulation characteristics and the associated rights and obligations for companies and investors alike, security tokens need to be considered on their own. The existing literature on STOs studies success determinants during the funding process regarding investors' rights, issuer, and offering characteristics (Lambert et al., 2022) or cheap human capital and social media signals (Ante and Fiedler, 2020). Momtaz (2023) decribes the economics, law, and technology of STOs and provides a comparison of STOs, ICOs, and IEOs. Other studies embed STOs in a theoretical context, e.g., Gan et al. (2021) study the optimal design of an STO, Gryglewicz et al. (2021) examine when token financing is preferable to equity financing, while Miglo (2021) compares STOs and ICOs under moral hazard and demand uncertainty.

We theoretically embed this article in the context of signaling theory to overcome information asymmetries between the STO-issuing company and potential primary and secondary market investors both during the pre-and post-STO phase. This article extends previous research by investigating whether a pre-sale and the announcement of token transferability or later expected liquidity is positively related to the success of an STO. They can be interpreted as positive quality signals and have not been investigated in the context of an STO yet. During a pre-sale, the transparent investment of publicly known experts and institutions serves as a signal for trustworthiness (Howell et al., 2020) and constitutes a method to gather valuation-relevant information at an early point of the process to make the main funding more effective (Momtaz, 2020). We find that a pre-sale and the announcement of transferability serve as quality signals, and both have a positive link to the funding success of an STO. The announcement of future token transferability enables the investors to trade the tokens on secondary marketplaces and should translate into liquidity in the post-STO phase. Once trading begins, the market valuation should lead to accurate pricing and show whether the signals previously sent about the quality of the STO correspond to reality (Florysiak and Schandlbauer, 2022). In this regard, to the best of our knowledge, we are the first to empirically investigate the post-STO phase by analyzing the secondary market for security tokens. As the first market valuation, we study underpricing and relate it to the literature on IPOs regarding determinants such as market sentiment and large investors. Underpricing hardly seems to exist in the STO market, whereby it is related to the market sentiment as an external signal. As a further market valuation, we examine the short-term post-listing performance by calculating buy-and-hold as well as buy-and-hold abnormal returns over different short-term horizons. In this way, we can verify whether the signals previously sent reflect the reality of the STO and potentially translate into higher returns. We find that both extremely negative and positive returns can be achieved depending on the time horizon. Furthermore, we analyze the evolution of the liquidity situation in the market since its inception. In particular, we add to the literature the substream of research that disentangles the effect of a token being traded on centralized or decentralized exchanges as a means of the rapidly evolving area of decentralized finance. So far, this has solely been elaborated for cryptocurrencies as a whole by Aspris et al. (2021) but has not been addressed in any other previous study on the aftermarket performance of tokens. Our study is based on two hand-collected, overlapping, but non-identical datasets comprising 138 STOs and 108 security tokens traded on the secondary market.

The remainder of this paper is organized as follows. In Section 3.2, we present the technological background and classification of STOs. In Section 3.3, we present an overview of signaling theory and derive our hypotheses. Section 3.4 describes data, variables, and results regarding the pre-STO phase and the analysis of STO success determinants. Section 3.5 focuses on the post-STO phase, including STO underpricing, returns to investors, and liquidity. Section 3.6 concludes this study.

3.2 Security token offerings: Background

3.2.1 Technological background

We first describe the technological background and termini relevant in the context of a security token offering (STO) on the distributed ledger technology (DLT). DLT refers to an approach in which data is recorded and shared via a decentralized, distributed ledger of various different participants. The blockchain is the most relevant form and sub-category of DLT, although both terms technically are not identical (Fisch, 2019), however, we use the terms synonymously in this study. The structure in the form of cryptographic chains of data blocks is characteristic of a blockchain. Anyone can see and download a copy of the blockchain. The only relevant version is the one that contains the latest legitimate transactions (Schär and Berentsen, 2020). The immutability of the blockchain and its transactions generate trust between the parties involved (Chod et al., 2022). Ethereum is the most commonly used blockchain infrastructure for ICOs (Howell et al., 2020) as well as for STOs. This has prevailed due to the wide range of application possibilities regarding the programming and execution of smart contracts. Smart contracts are digital contracts that allow specific transactions to be executed automatically when certain predefined events occur (Buterin, 2013). The addition of assets to the blockchain is referred to as tokenization, while the digital version of the asset on the blockchain is called a token (Schär, 2021). The financial use case for smart contracts is these digital tokens, where the smart contract verifies, for example, that the investor has received payment and then automatically sends the token to the investor's wallet (Cong et al., 2022). The distinction between the three following types of tokens has crystallized (Howell et al., 2020), though there are several hybrid forms. Payment tokens are a means of payment for the purchase of goods or services (e.g. Bitcoin), utility tokens entail consumptive rights to use blockchain-based services and security tokens. For security tokens, we apply the definition of Lambert et al. (2022) as "a digital representation of an investment product, recorded on a distributed ledger, subject to regulation under securities laws" (Lambert et al., 2022, p. 302). The application of blockchain to the entire financial sector holds great potential for systemic change (Guo and Liang, 2016; Wright and De Filippi, 2015).

3.2.2 Implications for financial markets

The tokenization of assets has multiple implications for investors, companies, and financial markets alike. The global nature of the blockchain, and thus the lower barriers between financial markets of different countries, means companies have a wider geographic scope and can reach a broader investor base (Chang, 2020). Fractional ownership through the divisibility of the underlying asset enables retail investors to invest small amounts of money in previously unattainable asset classes, which allows investors to diversify their portfolios more broadly. For tokenized assets, investors no longer need to demand higher returns resulting from higher divestment risk. Therefore, tokenized assets can reduce illiquidity premia and finally make these assets trade closer to their fair value (OECD, 2020). The properties of the blockchain promise increased transparency in tamper-proof, instantaneous transactions. Automated transaction processing, as well as the allocation and distribution of payment flows using smart contracts, can reduce the costs of issuance and transactions (Chang, 2020; Guo and Liang, 2016). Automated settlement and disintermediation lead to a reduction in trading fees and a significant decrease in settlement times, thereby enabling more efficient financial markets (Momtaz, 2023). Moreover, by leveraging a blockchain, tokenization can eliminate the counterparty risk since intermediaries become obsolete (Uzsoki, 2019). All of these technical innovations are paving the way for a digitized token economy of the future.

3.2.3 Differentiation from existing forms of financing

IPOs are the traditional, regulation-compliant way to list a company publicly for the first time. A common feature of IPOs and STOs is that the offering has to comply with regulations, and investors receive binding rights. A substantial difference between STOs compared to IPOs is the use of a blockchain. This ensures that the settlement of the transactions after an STO is faster and more efficient (Mills et al., 2016). The issuance and marketing processes of IPOs and blockchain-based offerings are completely different: IPOs perform a book-building process and use social media solely to attract investors; token offerings communicate relevant financing information for the offering to prospective investors through social media channels (Ofir and Sadeh, 2020).

The basic idea behind crowdfunding (CF) is that funding of a target amount is achieved by collecting small amounts of money from the crowd of investors – this is a common feature with STOs due to the division into tokens. For CF, platforms handle the projects holistically, act as intermediaries, and perceive monitoring functions in the selection process of the projects. In ICOs or STOs, platforms play only a subordinate role in displaying aggregated information about projects due to the decentralized blockchain, leading to a shift in screening activities exclusively to individual investors (Block et al., 2020). The problem with CF is that the shares purchased may be difficult to resell or liquidate because there is no real secondary market, while tokens can usually be traded on secondary markets.

Both CFs and ICOs are about raising money from potential users to spend later on the platform for services, outside of which the token has no value (Howell et al., 2020). Thus, utility tokens are legally classified only as donations with limited rights, while investors in regulated security tokens receive corresponding rights from the underlying financial instrument (Ante and Fiedler, 2020).

3.3 Theory and Hypotheses

3.3.1 Signaling theory

The conceptual framework of our hypotheses draws upon the literature in the field of information asymmetries and signaling. Signaling theory deals with reducing information asymmetries between the involved parties (Spence, 2002). In the case of STOs, these information asymmetries arise from the fact that the STO-issuing firm has internal, private information about its quality and future prospects that is not available to the public. The signal itself needs to be observable for the receiver and associated with monetary, time, reputation, or effort-related costs that prevent imitation (Connelly et al., 2011). Therefore, companies have an incentive to communicate this information to potential investors and reduce information asymmetries. As a result, investors are better able to identify high-quality ventures and invest accordingly (Bergh et al., 2014; Florysiak and Schandlbauer, 2022). Information asymmetries are especially prevalent in token offerings, as these companies are often young, and lack a solid track record and experience (Howell et al., 2020). This effect is amplified by retail investors, who are mainly present in the market for token offerings (Lee et al., 2022). In comparison to institutional investors, retail investors have less experience and financial resources to evaluate investment opportunities properly (Ahlers et al., 2015). Additionally, the underlying blockchain requires investors to have a certain level of technical knowledge and familiarity (Momtaz, 2021a). Consequently, it is crucial for companies conducting an STO to send quality signals to potential investors in order to reduce information asymmetries. Information asymmetries and the related signaling play an important role both during the STO on the primary market (the pre-STO phase) and when security tokens are traded in the secondary market (the post-STO phase).

3.3.2 Hypotheses Development: Pre-STO phase

An STO consists of several rounds, and a pre-sale can precede the actual main public offer. A pre-sale commonly aims at a limited group of investors and has several advantages. On the one hand, Howell et al. (2020) compare a pre-sale to the book-building process in IPOs to ascertain information about the correct demand and price, which makes the main funding more effective (Momtaz, 2020). Usually, a pre-sale has a discount on the token price for early investors. A pre-sale could therefore lead to early participation and a momentum effect (Roosenboom et al., 2020) due to the authentication of the issuer, especially when prominent experts or institutions can be attracted (Howell et al., 2020). In the context of reward-based and equity crowdfunding, it is found that the generation of early investors and an early, strong campaign is a quality signal of project success for potential investors (Colombo et al., 2015; Vulkan et al., 2016). The possibility to costly gather price-relevant information and attracting early attention before the main offering could signal that the STO is of high quality, which may be perceived as positive by investors.

> **Hypothesis 1:** *The implementation of a pre-sale phase is positively related to the success of an STO.*

There are two ways to trade and transfer a security token: on exchange platforms or directly from peer-to-peer (P2P). Even if a security token is not listed on an exchange platform, an investor can generate liquidity via a P2P transaction. The transferability of the token is constitutive of the possibility of obtaining future liquidity by trading the security token. From a technical standpoint, the feature of transferability of a token cannot be taken for granted. Some companies point out that the issued token may not be transferred and that the transferability will therefore be technically restricted over the course of programming the token.[1] This technical limitation restricts the future liquidity of the token. Florysiak

[1] Vermögensanlagen-Informationsblatt RAAY Real Estate GmbH, 2020: "Investors do not have the right to transfer and encumber the token to third parties. An obligation of the issuer or the company to take back the token exists through the right of termination.[...] A sale of the token by the investor is generally not possible." [translation by the authors]

and Schandlbauer (2022) even go so far as to claim that a security token gets its value from the fact that it is tradable. Already in the ICO context, it is stated that technical aspects of the technology used, such as the transferability of the token, play a major role in the investment decision of an investor (Fisch et al., 2021). Transferability is a major advantage of STOs over crowdfunding. For equity crowdfunding, a platform is explicitly required to trade the shares due to the lack of a blockchain (Signori and Vismara, 2018). Investors could therefore rate the announcement of transferability of the security token as a quality signal and invest primarily in STOs in which they can resell the security token without restrictions from the issuing company to generate future liquidity. The explicit emphasis on the intent of transferability is a potential indicator of high-quality STOs and shows that they intend to trade their tokens in the secondary market in the future, thus deriving value.

> **Hypothesis 2:** *The announcement of transferability is positively related to the success of an STO.*

Transferability is both a quality signal during the pre-STO phase and a technical prerequisite for tokens to be traded on the secondary market in the post-STO phase. The market valuation in the post-STO phase can be used to verify to what extent the signals sent during the STO correspond to reality and are subsequently reflected in the associated STO market outcomes.

3.3.3 Hypotheses Development: Post-STO phase

In the following, we focus on the post-STO phase by investigating underpricing or, more specifically, 'money left on the table' for the issuer (Loughran and Ritter, 2002). We account for underpricing as the return of an STO investor on the primary market who holds the token until the listing on the secondary market. We derive hypotheses for the determinants of STO underpricing that relate to external signals, in other words, the signals that come from outside the STO-issuing firm as opposed to the primary offering.

In the *increased monitoring hypothesis*, Stoughton and Zechner (1998) state that underpricing is a way to attract large investors under the assumption that only these investors are capable of monitoring. In practice, companies seek to incorporate large investors into the shareholder structure who have mechanisms to monitor and influence management in order to increase the firm value in the interests of all shareholders (Admati et al., 1994). Stoughton and Zechner (1998) state that small investors free-ride on large investors' monitoring, an agency-problem which is also documented in the context of equity crowdfunding (Hornuf and Schwienbacher, 2018; Moritz et al., 2015). Therefore to increase the firm value, the company needs to lure large investors with the help of underpricing in their own interest. As a consequence, the fewer large investors invested in the STO, the primary offering, the more pronounced the underpricing will be to incentivize large investors to invest in the secondary market.

Hypothesis 3: *The number of large investors during the STO is negatively related to underpricing.*

The IPO literature suggests that market sentiment is an important predictor of underpricing (Loughran and Ritter, 2002; Green and Hwang, 2012). The demand of sentiment investors may disappear in times of negative market sentiment, and, therefore, 'normal' investors with IPO stocks in inventory need to be compensated through underpricing for the associated risk of losses (Ljungqvist et al., 2006). We expect that the market for security tokens is salient to this kind of market timing since Baker and Wurgler (2006) have shown that investor sentiment is particularly present for subjective and difficult-to-arbitrage securities, such as security tokens. It is up to the STO-issuing firm when exactly the trading of their tokens on the secondary markets starts. In order to prevent their token from generating negative initial returns, they will time the first trading day and avoid phases of negative market sentiment (Drobetz et al., 2019). Consequently, we assume that issuers wait times of positive market sentiment and avoid negative market sentiment as an external signal, which increases underpricing.

Hypothesis 4: *The market sentiment is positively related to underpricing.*

3.4 Pre-STO phase

3.4.1 Sample construction and data of STO success determinants analysis

There is no central database of all STOs carried out to date. As such, this sample is obtained by manually collecting and matching data from multiple data sources and websites. First, the starting point was the website *Digital Asset Network*. From there, we moved to various aggregator sites and looked for offers declared as STO.[2] In the second step, we searched the companies' websites for information about each STO. For STOs issued in the USA, we additionally accessed the *EDGAR* database from the SEC. We collected documents such as white papers, legal documents, prospectus, and further investor documents. Third, a plausibility check took place to verify the collected data, including matching with transaction data from the blockchain, as information in different databases

[2]The aggregator websites considered in this study are Block Databank, Blockdata, BlockState, CoinMarketPlus, Digital Asset Network, ICO Bench, ICO Drops, ICO Holder, ICO Stamp, ICOs Bull, STO Analytics, STO Docket, STO Filter, STO Market, STO Rating, STO Scope, The Tokenizer.

may converge. The final step for each observation was to check the accordance with the definition of a security token of Lambert et al. (2022). We had to exclude many STOs due to limited data availability, STOs that were announced but for which there was never an offer and offerings that did not meet the definition. We executed these steps in sequence and obtain 71 STOs with very detailed data. We validated and complemented our self-collected data with 67 STOs from the *Token Offerings Research Database* (TORD) of Momtaz (2021b) after removing duplicates and follow-up research. Finally, we end up with 138 STOs. These STOs were issued between 1st March 2017 and 31st December 2020. The sample size in other STO success determinants studies is similar, especially the reduction due to missing detailed information to perform the multivariate analysis (Ante and Fiedler, 2020; Lambert et al., 2022).

3.4.2 Variables of STO success determinants analysis

The choice of the dependent variable as a measure of STO success is not completely trivial. In pure equity markets, naturally, a valuation-based measure is preferable, which relates the amount raised to the portion of equity sold by the issuer. For instance, two companies may raise the same amount of money in the STO but give up a different proportion of equity, resulting in different valuations. However, in addition to stocks, our sample also includes other financial instruments such as fund or debt tokens, whose observations we do not want to lose by opting for a valuation-based variable since the focus of this study is on new entrepreneurial funding mechanisms in general and not on the type of capital. Therefore, the *Funding Amount* serves as our simplified dependent variable reflecting a firm's overall ability to raise funds from investors and is thus the most direct way to gauge a firm's access to external finance (An et al., 2019). The use of the variable to quantify the success of a project is common in the literature on venture capital (Baum and Silverman, 2004), crowdfunding (Block et al., 2018; Mollick, 2014), ICOs (Fisch, 2019; Lyandres et al., 2022), and STOs (Ante and Fiedler, 2020; Lambert et al., 2022). Accordingly, our results need to be interpreted more from the investors' perspective, as they reflect the collective reaction of investors to the STO rather than the financial corporate valuation or implications thereof. To account for the high skewness of the *Funding amount*, we use a log transformation. As an alternative measure of success, we incorporate the variable *Funding amount to target* as an additional dependent variable. It is the percentage ratio of the *Funding amount* to the *Hardcap*, the pre-defined target amount of the STO. Considering this ratio allows us to address the issue that a few STOs with large *Funding amounts* may bias our results (Lambert et al., 2022).

To test Hypothesis 1, we include the dummy-variable *Pre-sale*, which takes a value of 1 if a firm conducts a *Pre-sale* phase before the main funding, and 0 otherwise. To test Hypothesis 2, we consider the dummy-variable *Transferability*, which accounts for whether a company announces in its published documents for the STO that the token is technically equipped to be transferable for investors.[3]

We include several control variables in our models. We control for different types and rights of tokens representing their economic purpose: *Equity token*, *Fund token*, and the remaining investment tokens. *Equity tokens* usually entail the investor with cashflows in the form of dividend payments. *Fund Tokens* are issued as security tokens that offer diversification opportunities through indirect investments, which makes them potentially attractive to investors. Additionally, the dummy-variable *Voting rights* refers to the possibility of the investor to participate, e.g., in the composition of the board or in structural decisions that provide the investor with opportunities for control. If STO investors are not entitled to a *Voting right*, it would be indicative of the typical corporate governance issue of separation between control rights and ownership (Lambert et al., 2022). We further control for several variables which are known from CF and ICOs. The dummy-variable *Softcap use* indicates whether a minimum funding threshold must be reached for an STO to be issued. The metric variable *Hardcap* measures the STOs' funding target for which a log transformation is used to account for the skewness. Investors have the incentive to select projects with realistic *Hardcaps*. A target amount set too high could indicate that the project will not reach the amount. A target amount that is too low could suggest that a project will not be carried out (Mollick, 2014) or that a campaign will stop early (Fisch, 2019). The variable *Telegram* describes whether a company makes use of Telegram as a communication medium. *Telegram* has established itself as a communication channel in the crypto world to communicate information directly with potential investors. The use of *Telegram* signals a company's familiarity with the general framework in the crypto sphere (Amsden and Schweizer, 2019). We additionally include variables related to the characteristics of the issuing company, as investors draw inferences about the quality of the offering from the firm. The variable *Listing* indicates whether an STO-executing firm is listed on a traditional stock exchange, which is a signal for potential maturity and regulation compliance of the company. We additionally control for the logarithmized *Age* of the company as the difference between the date of STO and the date of formation of the firm. The probability of a company's survival decreases more significantly in earlier years (Pazos, 2019). Investors could anticipate this and invest in older companies. Already in the crowdfunding context, the influence of geography on campaign success was identified (Mollick, 2014). Because of this, additional dummy-variables for the country of incorporation are included: *USA*, *Cayman Islands*, *UK*, *Europe*, and the remaining countries.

[3]While an investor could also glean this information from the smart contract, it cannot be assumed that the average investor has these technical capabilities. Therefore, we rely on the information provided in the offering documents.

3.4.3 Descriptive statistics of STO success determinants analysis

We report the descriptive statistics of the variables used in the analysis for STO success determinants in Table 3.1.

Table 3.1: Descriptive statistics for STO success determinants.

Statistic	N	Mean	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Dependent variables								
Funding amount	138	9.698	6.917	0.000	0.000	13.220	14.871	18.713
Funding amount to target	71	0.311	0.401	0.000	0.0001	0.120	0.524	1.070
Independent variables								
Pre-sale	138	0.377	0.486	0.000	0.000	0.000	1.000	1.000
Transferability	71	0.831	0.377	0.000	1.000	1.000	1.000	1.000
Equity token	71	0.366	0.485	0.000	0.000	0.000	1.000	1.000
Fund token	71	0.113	0.318	0.000	0.000	0.000	0.000	1.000
Voting rights	71	0.183	0.390	0.000	0.000	0.000	0.000	1.000
Softcap use	71	0.662	0.476	0.000	0.000	1.000	1.000	1.000
Hardcap	71	8.433	8.328	0.000	0.000	13.883	16.660	20.723
Telegram	71	0.563	0.499	0.000	0.000	1.000	1.000	1.000
Listing	71	0.056	0.232	0.000	0.000	0.000	0.000	1.000
Age	71	0.557	0.722	0.000	0.000	0.164	0.895	3.088
Cayman Islands	138	0.051	0.220	0.000	0.000	0.000	0.000	1.000
Europe	138	0.297	0.459	0.000	0.000	0.000	1.000	1.000
UK	138	0.087	0.283	0.000	0.000	0.000	0.000	1.000
USA	138	0.312	0.465	0.000	0.000	0.000	1.000	1.000

Note: This table reports the descriptive statistics (mean, standard deviation, minimum, 25^{th} percentile, median, 75^{th} percentile, and maximum) for the full sample. The different number of observations of N=71 and N=138 is based on the fact that not all of the variables considered in our analysis are included in the *Token Offerings Research Database* of Momtaz (2021b). All variables are defined in Table A.1.

The *Funding amount* has a mean of 9.698 corresponding to $16,285. The minimum and 25th percentile with a value of 0 indicate that there are many unsuccessful offerings. The maximum value of 18.713, which corresponds to $133,953,060, demonstrates the high skewness. The mean and median of the alternative success variable *Funding amount to target* reveal that the majority of companies do not reach the *Hardcap*. *Pre-sales* were conducted on average of 37.7% of the ventures to offer their tokens prior to the main funding phase. The share of companies offering a *Pre-sale* is, in comparison to ICO studies, lower with 43% (Howell et al., 2020), 53% (Florysiak and Schandlbauer, 2022), or even 65% (Fisch, 2019). The *Transferability* feature of the token to ensure future liquidity was mentioned by 83.1% of the ventures in their offering documents.

Our control variables regarding token types and rights show that most STOs with 36.6% issue an *Equity token* entailing dividend payments and 11.3% a *Fund token* as an indirect investment. A share of 18.3% of the tokens provides a *Voting right* to the investor, which is an indication of the separation of control and voting rights. The issuers do not intend to give investors a say in the company matters, which is consistent with the findings of Lambert et al. (2022). The control variables related to modern forms of venture funding reveal that 66.2% of the companies make use of a *Softcap* as a financing threshold. The *Hardcap* with a median of 13.883 corresponding to $1,069,819 and a maximum of 20.723 which corresponds to $999,734,198, both of which are higher than the actual *Funding amount*, indicate that most companies fail to meet their pre-specified *Hardcap*. On average, the mass-market communication channel *Telegram* is used by 56.3% of companies to communicate directly with investors. Table A.2 in the Appendix displays the correlation coefficients for all variables related to the analysis of STO success. The variance inflation factors (VIF) are reported below the regression coefficients in Table 3.2. We have neither high correlations above 0.5 nor VIFs above a conservative threshold of 5. Thus, we assume that multicollinearity is no concern in our analysis.

3.4.4 Multivariate Analysis: STO success determinants

Table 3.2 presents the results of the tobit models with *Funding amount* as the dependent variable. We estimate a tobit specification as the dependent variable *Funding amount* is left-censored at zero since we account for unsuccessful funding with a value of zero. All specifications are estimated with heteroscedasticity-robust standard errors and year dummies. Model (1) includes the STOs of the *TORD* of Momtaz (2021b) resulting in 138 observations, while models (2) to (5) are reduced to the smaller sample of 71 observations with more detailed data because not all of the variables included in our analysis are in the *TORD* database. However, the coefficients continue to have the same signs and similar significances. The hypotheses-related variables are included interchangeably and step-wise in the models (2) to (4). In the full model (5), company-specific variables are also taken into account. The following explanations refer to the full model (5) with a Pseudo R^2 of 0.129. The relation of the number of observations to the number of variables in our models could be suspicious for overfitting. Therefore, we additionally calculate the Akaike Information Criterion (AIC). We find that our full model (5) has the lowest AIC value compared to the other models, thus, it is the best-fit model for our data.

Model (5) in Table 3.2 shows that conducting a *Pre-sale* is positively associated with the *Funding amount*. The occurrence of a *Pre-sale*, indicated by the dummy-variable with a value of 1, equals a c.p. increase of 16,204% in the *Funding amount*.[4] This result is important for STO-issuing companies since it emphasizes that the course for a successful STO can be set early on by planning the individual STO phases, including a *Pre-sale*. According to the rationale of signaling theory, conducting a *Pre-sale* involves effort and costs

[4] Since the dependent variable *Funding amount* is logarithmized, we have a log-level model. We, therefore, apply the Halvorsen and Palmquist (1980) correction for an exact interpretation of the economic significance, i.e. for *Pre-sale*: $100(e^{\beta_1} - 1)\% = 100(e^{5.094} - 1)\% = 16,204\%$.

Table 3.2: Tobit STO success determinants analysis.

	Dependent variable:				
	Funding amount				
	(1)	(2)	(3)	(4)	(5)
Pre-sale	3.100**	4.970**		5.538**	5.094**
	(1.563)	(2.320)		(2.155)	(2.216)
Transferability			4.232**	5.018***	5.858***
			(1.770)	(1.761)	(1.676)
Cayman Islands	9.720***	13.128***	11.192***	11.877***	12.839**
	(2.463)	(4.229)	(3.147)	(4.083)	(4.480)
Europe	6.589***	4.683	4.060	3.111	3.064
	(2.269)	(2.990)	(2.733)	(2.806)	(2.764)
UK	3.678	3.775	4.988	4.214	5.150
	(3.545)	(3.733)	(3.130)	(3.395)	(3.189)
USA	3.241	-0.023	0.294	-0.677	-0.893
	(2.335)	(3.016)	(2.684)	(2.726)	(2.839)
Equity token		2.019	1.697	1.784	1.714
		(1.699)	(1.690)	(1.596)	(1.487)
Fund token		3.123	3.316	1.650	1.848
		(3.327)	(3.358)	(3.076)	(3.172)
Voting rights		2.854*	3.787**	3.022*	3.383**
		(1.705)	(1.736)	(1.556)	(1.517)
Softcap use		−4.896***	−5.173***	−5.216***	−5.460***
		(1.357)	(1.326)	(1.279)	(1.223)
Hardcap		−0.288	0.022	-0.228	-0.025
		(0.483)	(0.469)	(0.455)	(0.450)
Telegram		−7.116***	−5.423***	−8.132***	−8.252***
		(2.113)	(1.675)	(1.914)	(1.900)
Listing					4.920
					(3.694)
Age					1.320
					(1.025)
Mean VIF	1.170	1.553	1.466	1.571	1.591
Maximum VIF	1.237	1.963	2.099	2.186	2.078
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	138	71	71	71	71
Pseudo R^2	0.047	0.101	0.095	0.116	0.129
Log pseudolikelihood	−373.305	−189.648	−190.956	−186.322	−183.715
AIC		409.236	411.912	404.645	403.431

Note: This table reports cross-sectional Tobit regressions. The reference category for the countries is *Country other*. All models include a not reported constant. Heteroscedasticity-robust standard errors in parentheses. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1.

for the STO and is therefore translated into higher signaling costs which only high-quality STOs can afford. Likewise, however, it is an easy-to-observe positive signal to potential investors that issuers are bearing these costs and are trying to gather valuation-relevant information to make the following main sale more effective. Consequently, we find empirical support for Hypothesis 1.

Moreover, the coefficient of *Transferability* is positive and significant. The announcement of *Transferability* in the offerings documents, indicated by the dummy-variable with a value of 1, equals a c.p. increase of 34,902% in the *Funding amount.* This finding underpins that the announcement of *Transferability* and the expectation of future liquidity enables companies to raise more funding. We find supportive evidence for Hypothesis 2 regarding the positive signaling effect of the announcement of *Transferability* to overcome information asymmetries. Interestingly, when embedding the results in the context of signaling, we observe that a company's intention to offer a transferable security token is crucially related to the success of an STO, even though it does not come at a high cost for the issuer and cannot be easily verified by investors. This may be due to the fact that the expectation to trade the token in the future appears to be the main motive for a token investment (Fisch et al., 2021).

The results pertaining to the token type and rights deliver only for *Voting rights* a positive and significant link to the *Funding amount.* This means that, unlike IPOs (Smart et al., 2008) and equity crowdfunding (Cumming et al., 2019), the separation of ownership and control does not play a major role for STOs. This result is in line with Lambert et al. (2022), who claim that the transparency of the blockchain and the associated lower costs of acquiring information for external investors reduce this agency problem. The coefficient of the variable *Softcap use* is negative and significant. Lambert et al. (2022) argue that if a *Softcap* is used, a company needs to convince more investors to reach the financing threshold in the first place. The utilization of *Telegram* as a communication channel to investors is negatively related to the success of an STO. Lyandres et al. (2022) claim that social media signals depend on the quality and cost of the social media platform, which is in the case of *Telegram* low. The *Cayman Islands* are positively associated with the *Funding amount.* However, we cannot disentangle the real considerations of the companies in this regard. On the one hand, the *Cayman Islands* are considered a tax haven with numerous tax advantages for investors, and on the other hand, they offer a more lax legal framework. For the remaining company-specific variables, we do not find a significant coefficient in any model specification.

As a robustness check displayed in Table 3.3, we estimate the tobit models with the alternative success measure *Funding amount to target* as the dependent variable.

In the alternative success specification, all signs remain unchanged, but the significance of *Pre-sale* disappears probably because of variation in our small sample (Lambert et al.,

Table 3.3: Robustness: Alternative success variable.

	Dependent variable:		
	Funding amount to target		
	(1)	(2)	(3)
Pre-sale	0.160		0.167
	(0.150)		(0.138)
Transferability		0.303***	0.406***
		(0.109)	(0.109)
Cayman Islands	0.665**	0.555*	0.655*
	(0.327)	(0.288)	(0.339)
Europe	0.235	0.167	0.135
	(0.166)	(0.153)	(0.169)
UK	0.095	0.126	0.176
	(0.191)	(0.169)	(0.178)
USA	−0.002	−0.008	−0.063
	(0.171)	(0.158)	(0.168)
Equity token	0.071	0.060	0.061
	(0.110)	(0.103)	(0.094)
Fund token	0.014	-0.010	-0.054
	(0.160)	(0.140)	(0.159)
Voting rights	0.194*	0.235**	0.244**
	(0.109)	(0.105)	(0.101)
Softcap use	−0.372***	−0.389***	−0.416***
	(0.105)	(0.097)	(0.092)
Hardcap	−0.066**	−0.054*	−0.048*
	(0.031)	(0.032)	(0.029)
Telegram	−0.363**	−0.326***	−0.443***
	(0.139)	(0.110)	(0.125)
Listing			0.398**
			(0.160)
Age			0.083
			(0.066)
Mean VIF	1.77	1.70	1.75
Maximum VIF	3.07	3.05	3.20
Year FE	Yes	Yes	Yes
Observations	71	71	71
Pseudo R^2	0.372	0.406	0.136
Log pseudolikelihood	−32.210	−30.438	−26.275
AIC	94.421	90.876	88.551

Note: This table reports the robustness checks for the STO success determinants analysis. Models (1) to (3) are tobit estimations with a left-censoring at zero with the alternative success variable *Funding amount to target* as the dependent variable. The reference category for the countries is *Country other*. All models include a not reported constant. Heteroscedasticity-robust standard errors in parentheses. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1.

2022). The coefficient for *Transferability* is still positive and significant, confirming our prior results. Interestingly, the company-specific variable *Listing* now loads positively and significantly, which is consistent with our expectation that this is an effective signal of a firm's maturity. We can conclude that the robustness check does not show major deviations from the main analysis.

There is a potential endogeneity issue with the explanatory variables *Pre-sale* and *Transferability* and the dependent variable *Funding amount.* An STO-issuing company might choose these features while there are some unobserved characteristics, such as the quality of the STO or the issuing company, that may affect both the choice of a *Pre-Sale* or *Transferability* of the issuer and the funding success. As a matter of fact, investors do not necessarily base their investment decision on *Pre-sale* and *Transferability*, but on other unobserved features. Consequently, we cannot completely rule out the possibility that our results are subject to an omitted variable bias.

3.5 Post-STO phase

3.5.1 Overview of ST secondary markets

The secondary marketplaces where security tokens can be traded are either centralized exchanges (CEX) or decentralized exchanges (DEX). Decentralized exchanges are one application case in the decentralized finance ecosystem and are marketplaces where transactions are performed through self-executing smart contracts without an intermediary. The key technical innovation of most DEX is a new model for liquidity provision called automated market making (AMM). While on a CEX, market-making works with conventional limit order books, and trades are settled on centralized servers *off-chain*, on a DEX it is automated *on-chain* via trading against a liquidity pool, a pool of tokens locked in a smart contract (Aoyagi, 2020).[5] Prices on a DEX are calculated automatically by an algorithm based on the liquidity that can be provided by anyone (Barbon and Ranaldo, 2022). Along with this, users of DEXs retain control over the private key of their token instead of transferring it to the exchange platform as in the case of CEX. Therefore, the tokens cannot be stolen during a hacker attack, ultimately lowering the counterparty risk (Lin, 2019). DEX can pave the way towards an 'on-ramping' of the tokens on a regulated CEX at a later point in time (Aspris et al., 2021). In the US, CEX need to be registered as Alternative Trading Systems (ATS); in Europe, they need an equivalent license as Multilateral Trading Facility (MTF), and they have to screen potential investors with respect to compliance to KYC and AML/CTF regulations.

[5]For a detailed description of the functioning of AMM and liquidity pools see Barbon and Ranaldo (2022), Lehar and Parlour (2022), Mohan (2022), and Schär (2021).

The choice of the marketplace by the STO issuer can be a signal of the quality of the security token. CEX screen the potential tokens to be listed and typically charge high listing fees as high entry barriers, which only high-quality companies with good future prospects can afford. In addition, CEX function similarly to traditional online marketplaces where investors do not need to be familiar with blockchain technology, making it easier to reach any investor. In contrast, DEX are not regulated, there is no listing fee, but requires familiarity with blockchain technology. Therefore, we assume that trading on a CEX, as opposed to trading on a DEX, is a signal for high-quality tokens and companies.

3.5.2 Data of STO Underpricing

Our first source for secondary market data is *stomarket.com*, and from there, we move to various exchange platforms.[6] The second data source are the blockchain explorers *ethplorer.io* and *etherscan.io* for information on the ownership structure. A concern of our dataset from the success determinants analysis is that only a minority of these security tokens are later listed on secondary markets.[7] This has multiple causes since we argued previously that not all projects intend to trade the tokens, and other projects are not successfully funded. The phenomenon of sample reduction is also commonly known in the ICO context (Fisch and Momtaz, 2020; Lyandres et al., 2022). Benedetti and Kostovetsky (2021) state that the majority of the money invested in ICOs is in tokens later listed on secondary markets. We complement the secondary market data for a holistic picture of the market by real estate STOs (RE STOs). We acknowledge that there may be some comparability issues between conventional and RE STOs. As in equity markets, though, the underlying business model is not as crucial to returns, liquidity, and related research questions, as REITs in indexes demonstrate. The RE STOs in our sample are not directly tokenized real estate, as this is currently difficult to implement from a legal perspective. As such, a special purpose vehicle is tokenized with the property as the only asset, and investors hold a deed to the cash flows of the company rather than acquiring ownership rights to the property. Additionally, the primary offering of RE STOs cannot be analyzed in the same multivariate setting as 'conventional STOs' in Section 3.4. The value of a property, based on the funding amount, is mainly determined by its property characteristics, such as size, location, or type of use. As such, information asymmetries during the primary offerings and signals to overcome them differ strongly. In addition, the inclusion of the real estate sector seems reasonable, as Howell et al. (2020) document that the success of token

[6] We consider the following CEX and DEX for security tokens in our analysis: tZERO, INX Securities, Tokensoft, Openfinance, CryptoSX, Securitize Markets, Uniswap, Levinswap, StellarX, and MERJ.

[7] Note that there is a difference between the *Transferability* analyzed in Section 3.4 and the listing on the secondary market. *Transferability* refers to the technical property that the programmer has allowed the tokens to be transferable after the issuance when programming the smart contract, which companies can disclose in the STO prospectus. Whether a company actually lists the tokens on the secondary market is an entirely different matter, for which *Transferability* is merely the technical prerequisite.

offerings is particularly pronounced when it comes to business models that involve the tokenization of real assets. Nevertheless, when the tokens enter the secondary market, the market dynamics close these information asymmetries, and the market valuation, as well as the trading behavior, are similar. In any type of STO, investors receive regular cash flows from their tokens, whether in the form of a dividend, coupon, or rent payment. As mentioned earlier, this study focuses on the technical aspects of new funding mechanisms on the blockchain, which is why we consider RE STOs as valuable additional observations. Our sample covers the period from January 1st, 2019, to 31st December 2021. The time difference of one year compared to the success determinants sample is due to the fact that many tokens are not immediately traded on secondary markets or are even legally ineligible because of lock-up periods, as in the USA.

3.5.3 Variables STO Underpricing

Our dependent variable in the following analysis is *Underpricing*, which we define as the return in Equation 3.1 between the price of the token in the STO $P_{i,0}$ and the first traceable price on the market $P_{i,1}$.[8]

$$Underpricing = \frac{1}{n}\sum_{i=1}^{n}\frac{P_{i,1} - P_{i,0}}{P_{i,0}} \tag{3.1}$$

In the IPO literature, underpricing is a well-known phenomenon for which a plethora of theories, periods, and results in multiple markets have been investigated over the years.[9] In the following, we transfer explanatory approaches from IPOs that appear relevant to the STO context, which may also share similarities with technology or "New Market" IPOs. First, we refer to the *market liquidity hypothesis* of Aggarwal et al. (2002), which suggests that companies pursuing a token offering face pressure to underprice to obtain market liquidity to signal their future growth potential. This, in turn, generates an information momentum, capturing the broad interest of media and analysts (Aggarwal et al., 2002), who can also fulfill certification tasks of the issuer (Booth and Smith, 1986). This liquidity enables the companies to avoid illiquidity premia, compensates early investors for the undertaken risk, and causes network effects (Momtaz, 2020). Consequently, the liquidity generated by underpricing is an opportunity for companies to attract investors (Brau and Fawcett, 2006), while it also serves to mitigate information asymmetries. Further, another related theory is that higher information asymmetries are associated with higher underpricing (Rock, 1986; Welch, 1989), which implies that potential high-risk IPOs are

[8] Contrary to our approach, other studies in the ICO context refer to as underpricing the first-day-return between the opening and closing price on the first trading day (Momtaz, 2020, 2021a), which we calculate separately in Section 3.5.6.

[9] For a literature review on underpricing, see Ljungqvist (2007).

more underpriced (Ritter, 1984). This phenomenon also applies to STOs, as most companies that carry out an STO cannot present a comprehensive track record, experience, or a market-ready product resulting in high information asymmetries.

To analyze the influence of different investors involved in STOs as outlined in Hypothesis 3, we include the *No. large investors* as a numerical count of the number of investors who hold a share of more than 5% of all issued tokens. We use the 5% threshold related to the Schedule 13D filing, a disclosure requirement to the SEC in the USA for investors who acquire more than 5% of the beneficial ownership of a company. We derive the ownership information from the blockchain explorers at the date of the token issuance.[10] To test Hypothesis 4, we consider the variable *Sentiment* as the 30-day return of Ether on the first day of trading. As stated in Section 3.2, Ethereum is the dominant blockchain platform for STOs, and therefore the return of the corresponding native token Ether is an appropriate benchmark for the underlying market sentiment. We derive the data from *Coinmarketcap*.

We further control for the *Public float* of the tokens, which represents the percentage of the issued tokens that is attributed to investors who hold a share of less than 5%. A higher share of *Public float* was found to increase liquidity on stock markets (Ding et al., 2016). We include the logarithm of the *Trading volume* during the first 24 hours of trading. This measure reflects the actual interest of investors in an STO, resulting in a movement to the true market price (Felix and von Eije, 2019). For IPOs, Schultz and Zaman (1994) provides empirical evidence that underpriced stocks are traded more often on the first trading day than fully-priced stocks. We consider the dummy-variable *DEX*, which equals 1 if the token is traded on a decentralized exchange and 0 if it is traded on a centralized exchange. To take into account the prior success in the STO as analyzed in Section 3.4, we consider the logarithms of the variables *Funding amount* and *Token price*. Furthermore, we include the dummy-variable *STO type*, which equals 1 for 'conventional STOs' and 0 for real estate STOs to control for potential differences regarding *Underpricing*.

3.5.4 Descriptive Statistics STO Underpricing

We present the descriptive statistics for the variables used in the STO underpricing analysis in Table 3.4.[11]

We winsorize *Underpricing* at the top and bottom 5% to account for extreme outliers. The average *Underpricing* amounts to 1.2% with a median value of -2.1%. This means that

[10] We only consider unique wallet addresses of investors and their shares. However, due to blockchain technology, we cannot further ascertain what kind of investor it is.

[11] Additional detailed descriptive statistics for the conventional STO and RE STO sub-samples are presented separately in Table A.3 in the Appendix. It can be observed that there is a disparity between the *Funding amount* and the *Token price* of conventional and RE STOs. However, since the dependent variable *Underpricing* is a fraction of prices, the absolute differences regarding higher *Funding amounts* or *Token prices* are thus scale-free.

Table 3.4: Descriptive statistics for STO Underpricing.

Statistic	N	Mean	SD	Min	Pctl(25)	Median	Pctl(75)	Max
Dependent variable:								
Underpricing	106	0.012	0.144	−0.156	−0.054	−0.021	0.007	0.490
Independent variables								
No. large investors	106	3.132	1.574	1.000	2.000	3.000	4.000	10.000
Sentiment	107	−0.001	0.236	−0.583	−0.113	−0.113	0.098	1.289
Public float	106	0.404	0.322	0.000	0.050	0.526	0.705	0.862
Trading volume	107	3.187	2.477	0.000	1.800	2.700	4.200	12.000
DEX	106	0.830	0.377	0.000	1.000	1.000	1.000	1.000
Funding amount	106	12.238	2.590	0.000	11.031	11.110	12.932	18.713
Token price	106	3.481	1.296	0.010	3.891	3.961	4.009	7.311
STO type	107	0.196	0.410	0.000	0.000	0.000	0.000	1.000

Note: This table reports the descriptive statistics (mean, standard deviation, minimum, 25$^{\text{th}}$ percentile, median, 75$^{\text{th}}$ percentile, and maximum) for the full sample. The variable *Underpricing* is winsorized at the top and bottom 5%. All variables are defined in Table A.1.

the average STO leaves money on the table, whereas the median indicates overpricing at the cost of the investors. Both the mean and the median values are not far from zero, implying that the majority of the tokens are correctly priced. The minimum of -15.6% and maximum of 49.0% show that there are also companies with extreme over-and underpricing. In general, there does not appear to be underpricing in the ST market, as young companies lack experience, and the market is still in its infancy. The various results for *Underpricing* in the "New Markets" across Europe and for ICOs are substantially higher (Adhami et al., 2018; Drobetz et al., 2019; Felix and von Eije, 2019; Giudici and Roosenboom, 2004; Kiss and Stehle, 2002). This may suggest that the "New Market" is technologically not comparable to STOs or that the discrepancies in *Underpricing* results over time are the cause. For ICOs, this is not surprising, as information asymmetries are much more pronounced in completely unregulated ICOs than in STOs. For STOs, the ventures have to issue regulation-compliant prospectus, while unaudited white papers in ICOs mainly present basic information (Florysiak and Schandlbauer, 2022).

In an average ST traded on secondary markets, 3.132 large investors are involved at the date of the issuance. The *Sentiment* shows that security tokens become listed on the secondary market on average during days of slightly negative or neutral sentiment represented with -0.1% of the 30-day Ether return, while the minimum of -58.4% and maximum of 128.9% demonstrate the great variation of crypto returns.[12] On average, a share of 40.4% of all security tokens is attributed to the *Public float.* The logarithm of the *Trading volume* during the first 24 hours of trading has a mean of 3.187 which represents $24.22. In

[12] Note that since the beginning of the observation period, the Ether price has increased from $141 in January 2019 to $3,683 in December 2021.

our sample, 83.0% of the security tokens are traded on a *DEX* and the remaining on a regulated CEX. We use logarithms for the variables *Funding amount* for which the average is 12.238, corresponding to $206,489, and the *Token price* with 3.481 which corresponds to $32.49. The *STO type* reveals that 19.6% of the STOs are 'conventional STOs' and the remaining real estate STOs. Table A.4 in the Appendix shows the correlation coefficients for all variables. Although there are occasional higher correlations between *DEX* and the *Funding amount* of -0.784 or the *Token Price* with -0.716, all other correlations are below 0.5. Therefore, we do not include these variables in the same model since they could bias the regression coefficients. We report the VIFs in Table 3.5, all of which are far below a conservative critical value of 5. Hence, multicollinearity is unlikely to be an issue in the subsequent analysis.

3.5.5 Multivariate Analysis: STO Underpricing

The regression estimations of the determinants of STO underpricing are reported in Table 3.5.

The signs of the coefficients are consistent across the model specifications, and the adjusted R^2 amounts in all models are around 35%. The coefficient of *No. large investors* is only in model (2) significant at the 10% level on *Underpricing.* It appears that the *increased monitoring hypothesis* does not apply to the STO context. A possible explanation for this could be that Stoughton and Zechner (1998) refer to IPOs and thus pure equity, although our sample also includes debt or funds with different pricing dynamics. Notably, this finding aligns with the counter-intuitive results of Franzke (2004), which suggest that VC-backed IPOs, to which increased monitoring activities are attributed, experience higher levels of underpricing in the German "New Market" compared to those without VC-backing. Summarizing, we cannot provide statistical support in favor of Hypothesis 3. We find a positive significant link between *Sentiment* and *Underpricing.* A one-standard-deviation increase in *Sentiment* is in the model (1) associated with a 36.38% increase in *Underpricing* relative to the average. The results indicate that the *Sentiment* increases *Underpricing*, and issuers seem to time the first trading of their tokens to periods of positive market sentiment, which serves as a positive external signal. Our findings are in line with the IPO (Ljungqvist et al., 2006) as well as the ICO (Felix and von Eije, 2019) literature. Thus the conjecture in Hypothesis 4 that the crypto market *Sentiment* has a positive influence on *Underpricing* is supported by our empirical evidence.

As for our control variables, the coefficient of *Trading volume* is across all model specifications positive on *Underpricing.* The results are in line with IPO (Zheng and Li, 2008) and ICO literature (Felix and von Eije, 2019). None of our model specifications yield a significant effect of *DEX* and *STO type*, which is why we cannot observe any significant

Table 3.5: Determinants of Underpricing.

	Dependent variable: Underpricing					
	OLS			*Heckman*		
	(1)	(2)	(3)	(4)	(5)	(6)
No. large investors	0.015	0.015*	0.011	0.014	0.012	0.010
	(0.010)	(0.008)	(0.008)	(0.010)	(0.008)	(0.009)
Sentiment	0.222***	0.227***	0.216***	0.241***	0.236***	0.218***
	(0.064)	(0.062)	(0.064)	(0.057)	(0.058)	(0.064)
Public float	0.002	0.010	0.019	−0.059	−0.034	−0.039
	(0.030)	(0.029)	(0.028)	(0.072)	(0.082)	(0.093)
Trading volume	0.015*	0.015*	0.016**	0.015*	0.014*	0.016**
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)
DEX	−0.082			−0.077		
	(0.060)			(0.054)		
STO type		0.087			0.087*	
		(0.054)			(0.051)	
Token price			−0.037**			−0.033**
			(0.015)			(0.016)
Mean VIF	1.142	1.136	1.095			
Maximum VIF	1.230	1.194	1.110			
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	105	105	105	254	254	254
Adjusted R^2	0.343	0.351	0.405			
Log Likelihood				−48.053	−47.319	−45.062
ρ				−0.426	−0.343	−0.385

Note: This table reports cross-sectional OLS regressions for the determinants of STO Underpricing in models (1) to (3). Models (4) to (6) present the results from the Heckman (1979) procedure using maximum likelihood estimation with the selection variable *Funding amount*. Heteroscedasticity-robust standard errors in parentheses. All models include a not reported constant. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1.

difference between tokens traded on centralized and decentralized exchanges or conventional and RE STOs regarding *Underpricing*.[13]

A major criticism could be that this sample potentially suffers from a selection bias resulting from issuers that offer the tokens with a larger discount during the initial offering in order to increase the chance of a subsequent listing. We address this issue similarly to Benedetti and Kostovetsky (2021) and Florysiak and Schandlbauer (2022) by applying the Heckman selection model (Heckman, 1976, 1979). We perform a full information maximum likelihood estimation with the selection variable *Funding amount* since this is the major variable of

[13]We would like to point out that the signs and significances are consistent across all model specifications, regardless of whether the *Funding amount* and *Token price* are included or not.

STO success (as outlined in Section 3.4) and crucial for a token to become listed. We can therefore address this sample selection problem in a methodologically appropriate way and consider all listed and unlisted STOs simultaneously, which increases the number of observations. The descriptive statistics for this sample are displayed in Table A.3 in Panel C in the Appendix. Models (4) to (6) in Table 3.5 display the results from the Heckman procedure and have consistent signs as the previous models. We observe that *No. large investors* is no longer significant, whereas the positive and significant influence of *Sentiment* on *Underpricing* remains. Interestingly, in the model (5), the *STO type* is significant on the 10% level on *Underpricing*, meaning that 'conventional STOs' have a higher *Underpricing* in comparison to real estate STOs. This could be due to the fact that the price of real estate can be more accurately determined and is more transparent to the public, making these STOs more likely to be priced correctly. We conclude that a potential selection bias is rather unlikely to be driving our results.

3.5.6 Returns to investors after the token listing

As a further market valuation, we validate the previously sent signals about the quality of the STO by examining secondary market returns. We analyze buy-and-hold returns (BHR) as well as buy-and-hold-abnormal returns (BHAR) of investors who buy the security tokens on the first day the token is traded on an exchange and hold the token for different short-term time horizons ranging from one day to one year. We concentrate on this approach, e.g., the common risk factor models of Fama and French (1993) and Carhart (1997) rely on a longer data history to calculate expected returns, which is not yet available for security tokens. We calculate the raw buy-and-hold return (BHR) in the same way as *Underpricing*, but from the first day of trading $t = 1$ to the last day of the holding period T.

Alternatively, to calculate the buy-and-hold abnormal return ($BHAR$), we adjust the raw return by a value-weighted market capitalization-based benchmark, similar as Fisch and Momtaz (2020) and Momtaz (2021a) as follows:

$$BHAR = \frac{1}{n}\sum_{i=1}^{m}\Big[\frac{P_{i,t=T} - P_{i,t=1}}{P_{i,t=1}} - \sum_{j=1,j\neq i}^{n}\frac{MC_{j,t=T}}{\sum_{j=1}^{n} MC_{j,t=T}} \cdot \frac{P_{j,t=T} - P_{j,t=1}}{P_{j,t=1}}\Big], \tag{3.2}$$

where $P_{i,t=1}$ is the price of the security token i at the end of the holding period T and $MC_{j,t}$ refers to the market capitalization of the security token j on day T ($i{\neq}j$). The market consists of all security tokens with available price data. The value-weighted market benchmark is the product of the raw return of every other security token j over the holding period T and the market capitalization of a security token j over the sum of the whole market capitalization at the end of the holding period T. The adjustment for the market capitalization is suitable for several reasons. Firstly, some small-cap firms experience extreme returns, which could cause severe distortions of the results when using, e.g. volume-weighted or equally-weighted benchmarks (Momtaz, 2020). Secondly, market capitalization is subject to boom-and-bust cycles in the entire token market (Chen et al., 2021), which we can take into account in this way. The results of the *BHR* and *BHAR* analysis are displayed in Table 3.6.

Table 3.6: Analysis of BHR and BHAR.

		BHR Mean (Median)	*BHAR* Mean (Median)	Volatility
1 Day	106	0.231	0.229	0.258
		(0.010*)	(0.002)	
1 Week	106	0.015	−0.105**	0.364
		(−0.030**)	(−0.088**)	
1 Month	105	0.037	−0.152**	0.318
		(−0.019*)	(−0.026*)	
2 Months	102	0.062*	−0.300	0.330
		(0.006*)	(0.020)	
3 Months	98	0.050	−0.364	0.324
		(0.006**)	(−0.077**)	
6 Months	87	0.092*	−0.342**	0.420
		(−0.005)	(−0.351**)	
1 Year	24	0.549	0.136	2.256
		(−0.011)	(0.047)	

Note: This table reports the raw buy-and-hold returns (BHR) and buy-and-hold abnormal returns (BHAR) adjusted by a value-weighted market capitalization-based benchmark over different short-time horizons ranging from one day to one year. The mean, in parentheses, the median, and the volatility are displayed. The symbols * and ** denote significance at the 5% and 1% levels, respectively. All variables are defined in Table A.1.

Both the *BHRs* and *BHARs* vary depending on the investment horizon. The number of tokens diminishes over time, as many tokens have been listed in the last year of the observation period, and others have no continuous trading history as they e.g., changed the exchange platform to increase liquidity. Similar to the results in the ICO literature, we document partly extreme high ratios of mean to the median that exemplify the highly skewed distribution of returns in the market for tokens (Momtaz, 2021a). Particularly the high negative mean *BHARs* for holding periods between one week of -10.5% to six months of -34.2% trace back to the current situation on ST secondary markets where a few tokens which suffered substantial decreases in value make up the majority of the market capitalization. On the one hand, this shows the high probability of losses and, on the other, provides further evidence for the rationale that investors need to be compensated for the high risk they take by investing in a company with a weak track record (Benedetti and Kostovetsky, 2021). These findings are consistent with Kiss and Stehle (2002), who observe a post-IPO underperformance in the "New Market" between 1997 and 2001. A naïve investor who invests the same amount of money in every security token experienced, e.g., for a holding period of six months, a positive BHR of 9.2%, indicating potential wealth gains. Nevertheless, the corresponding medians fluctuate around the zero point over any holding period. In contrast, considering market capitalization, a security token investor realizes partially extreme negative and positive mean values of the $BHAR$. The medians draw a similar picture. To conclude this section of the post-STO performance, we observe both extremely negative and positive BHR and $BHAR$ over different short-term investment horizons.

3.5.7 (Il-)liquidity on secondary ST markets

A key benefit and promise of digital tokens is liquidity due to reduced costs and faster settlement times on the blockchain (Yermack, 2017), particularly because of the new method of liquidity provision. We investigate the liquidity situation on the ST secondary market since its inception, as liquidity is central for future industry development. In Figure 3.1, we display the development over time of several key characteristics of ST secondary markets.

The *Market capitalization* shows a strong positive trend, with stagnation in 2019 and during the beginning of the covid-19 pandemic, followed by a strong growth trend. A similarly positive growth trend is evident for the daily *Trading volume*. The high variability of the daily *Trading volume* relies on the fact that CEXs partly have trading hours just like conventional trading platforms and DEX operate continuously.

The liquidity situation on the market can be explained by the 'chicken-and-egg' problem, at least in the beginning when mainly CEX operated. On the one hand, investors expect to trade many different qualitative tokens while issuers will only pay the listing fees of the exchanges if the latter provides liquidity (Lambert et al., 2022). The analysis of the liquidity on cryptocurrency markets faces the problem of lacking high-frequency intraday data to determine high-frequency bid-ask spreads (Brauneis et al., 2021). As such, other metrics addressing the issue of low-frequency liquidity markets have to be considered. Firstly, we calculate the *CS estimator* of Corwin and Schultz (2012) as a simple bid-ask spread from daily high and low prices; see the detailed formula in the Appendix. Secondly, we compute *Liquidity* based on a modified version of the illiquidity measure of Amihud (2002) and Amihud et al. (2006), which originally determines the trading volume required

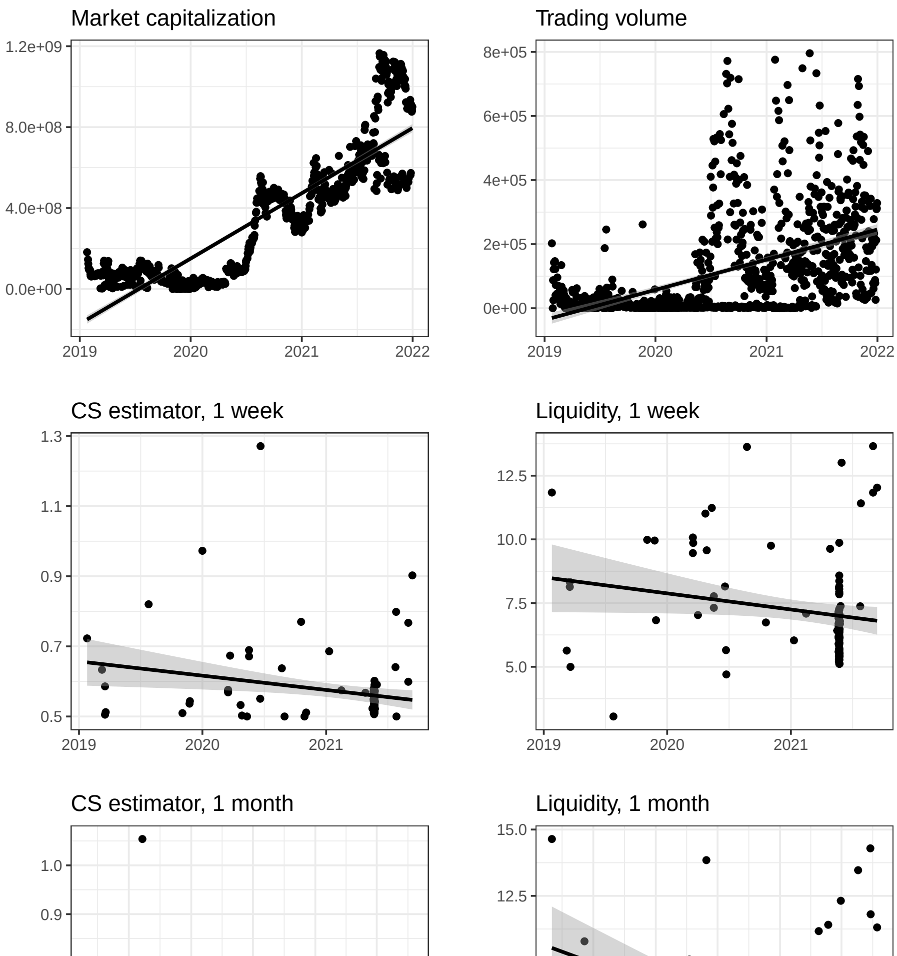


Figure 3.1: This figure presents the evolution of the security token secondary market from 1st January 2019 until 31st December 2021. The black line is the best-fit line. The *Trading volume* is censored at $800,000 because of the scaling. The variables are defined in Table A.1. N=108.

to move the price by 1%, as follows:

$$Liquidity_t = -log\frac{1}{5}\left[\sum_{t=t-5}^{t}\frac{|log(\frac{p_t}{p_{t-1}})|}{p_t \cdot volume_t}\right], \tag{3.3}$$

where it is multiplied by the negative of the logarithm to facilitate the numerical interpretation (Howell et al., 2018; Lyandres et al., 2019).[14] We consider both measures over an observation period of one week and one month after the first trading day and average them over five days. Figure 3.1 reveals that a large number of tokens were newly listed in 2021, which are mainly tokens on DEXs, as DeFi experienced tremendous growth in 2021.[15] The decrease of the *CS estimator* in Figure 3.1 over time indicates that the spread diminished, which is indicative of a more liquid market. Contrary, our *Liquidity* measure decreased over time, suggesting that especially newly issued tokens are less liquid. Brauneis et al. (2021) point out that, when studying liquidity levels, the Amihud et al. (2006) measure taking into account the *Trading volume* outperforms and is more meaningful than the *CS estimator*. Therefore, we conclude a general decreasing trend in liquidity on security token secondary markets over time. The graphical findings are empirically extended in the following. The calculation of our metrics with a sample split in CEX and DEX with a corresponding Welch t-Test for differences in mean (Welch, 1947) and the Mood Median-Test for differences in the median (Mood, 1950) are reported in Table 3.7.

The mean (median) *CS estimator* after a trading period of one week amounts for a CEX 0.64 (0.59) and for a DEX 0.56 (0.54), whereas after one month, it is 0.58 (0.54) and 0.53 (0.52). A direct comparison of centralized and decentralized exchanges is not possible as the differences in mean, and median are not significant. The mean (median) values of the *Liquidity* measure for a trading period of one week is on a CEX with 9.27 (9.66) and substantially lower for a DEX with 6.76 (6.45). The differences in mean and median are statistically significant, which underpins that decentralized exchanges are less liquid than centralized ones. For a trading period of one month, this finding is confirmed in the same way with an average (median) *Liquidity* on a CEX with 10.82 (9.66) and on a DEX with 6.67 (6.45) and highly significant differences in mean and median. These results are in line with Aspris et al. (2021), who find that CEXs are more liquid and that these tokens have a higher market capitalization which implies market segmentation and a reduction of governance risk. Hasbrouck et al. (2022) propose an increase in trading fees in an economic model as a solution to the low trading volumes on DEX. Both the *CS estimator* and the *Liquidity* measure reflect an increase in liquidity for a prolongation of the trading period from one week to one month. This fact is not surprising as trading activity can be limited

[14] The liquidity analysis is only included in the working paper version in Howell et al. (2018).

[15] We account for the increase of observations in 2021 in the empirical analysis with year-fixed effects in the underpricing regression models in Table 3.5, and we additionally verified the results in unreported analysis with a sample split and found no changes in our results.

Table 3.7: Security Token market characteristics.

		Exchange type		Tests	
		CEX	DEX	Mean Diff.	Median Diff.
CS estimator, 1 week	Mean	0.64	0.56	$t = 2.14^{**}$	
	Median	0.59	0.54		$X^2 = 0.67$
	SD	0.15	0.09		
CS estimator, 1 month	Mean	0.58	0.53	$t = 1.47$	
	Median	0.54	0.52		$X^2 = 1.12$
	SD	0.13	0.04		
Liquidity, 1 week	Mean	9.27	6.76	$t = 2.93^{***}$	
	Median	9.66	6.45		$X^2 = 3.62^{*}$
	SD	3.39	1.33		
Liquidity, 1 month	Mean	10.82	6.31	$t = 5.36^{***}$	
	Median	11.24	5.92		$X^2 = 6.73^{***}$
	SD	3.09	1.45		
	N	18	89		

Note: This table reports the mean, median, and SD (standard deviation) for the *CS estimator* and *Liquidity* after a trading period of one week and one month averaged over the last five days. The sample is split into centralized exchanges (CEX) and decentralized exchanges (DEX), for which the corresponding differences in mean are tested with a Welch *t*-Test and differences in the median with a Mood Median Test. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The variables are defined in Table A.1.

in the first trading days because the exact start of trading is not communicated beforehand, and investors on a CEX have to transfer their tokens to the platforms first before they start trading (Chanson et al., 2018). As comparative values to our results in terms of *Liquidity*, we consider utility tokens from ICOs with a mean value of 12.59 and NASDAQ shares as an industry benchmark with a much higher value of 18.16 (Howell et al., 2018). This comparison reveals that the security tokens in our sample are much less liquid than other investment possibilities so far.

Overall, it may seem as if liquidity has deteriorated over time, and the situation on security token secondary markets has worsened. However, more tokens have been listed primarily on less liquid DEXs over time. This is an indication of the slow completion of the range and the maturation of the market, which is driven by the increasing adoption of DEXs. For the tokens with low liquidity, it would have otherwise been unlikely to become listed on the secondary market at all. In this case, DEXs offer a simple way for a listing with low entry barriers and perspectives for a (cross-)listing on a CEX in the future, which so far is mainly used by high-quality security tokens. Meanwhile, the main problem is no longer the infrastructure but the lack of liquidity, which manifests itself in technology and global regulatory uncertainty as well as security concerns – in sum: trust and confidence in the system.

3.6 Conclusion

Security token offerings are a means for companies to raise capital where they issue digital tokens as regulated investment products on the blockchain. In this paper, we examine how signaling affects the market participants in the primary and secondary markets for security tokens, such as the STO-issuing company or investors in the primary and secondary markets. In order to obtain a holistic picture of the signaling effect on the entire market, we analyze STO market outcomes in the pre-STO phase and in the post-STO phase. We study success determinants of STOs which are a way for issuers to signal their quality to investors to overcome information asymmetries during the primary offering in the pre-STO phase. We find that both the execution of a pre-sale phase as a method to gather price-relevant information prior to the main funding and the announcement of token transferability as the expectation of future liquidity are positively linked to the funding success. In the post-STO phase, we find evidence that security tokens are hardly underpriced but are almost correctly priced with a mean (median) of 1.2% (-2.1%), indicating that issuers do not use underpricing as a way to attract investors. Drawing on the literature on IPOs, we show that underpricing is positively related to the sentiment on the crypto market, which serves as a positive external signal, and companies time the first notation of their tokens to avoid phases of negative market sentiment. Finally, the market valuation should reveal the true quality of security tokens. We find that over various short time horizons, both extremely positive and negative buy-and-hold (abnormal) returns can be achieved by an investor. Moreover, we conclude that the security token market lacks professionalism in investment evaluation and selection, as a naïve diversification strategy is a more promising approach to achieving high returns. We find that liquidity after the start of trading has decreased since the inception of the secondary market. However, this finding relies on the increasing number of tokens on less liquid decentralized exchanges. Decentralized unregulated exchanges offer lower entry barriers and complete the supply on the secondary market.

Our results highlight that companies that intend to raise funding via STOs would be well advised to offer a pre-sale phase in their STO and assure their intentions to trade the tokens on the secondary market while already devising a plan for successful future trading. From an investor's perspective, these signals can be interpreted as positive quality signals on the basis of which appropriate investment decisions can be conducted. Nonetheless, since extremely negative returns can also be achieved in the short term and there seems to be a lack of liquidity in the secondary market, investors should be well versed in the technical fundamentals and risks of blockchain investments. At this point, the legislator could also exert influence without at the same time over-regulating and hindering the further growth of the industry.

Our study has limitations. Because of the exclusion of several STOs due to limited data availability and the hand-collection of the data, we cannot completely rule out the possibility that a potential selection bias is present in our data. Therefore, the generalization and external validity of our results is reduced. Nevertheless, we collected and cross-checked data from various sources, such as the companies' websites, LinkedIn-Pages, aggregator websites, white papers, regulated prospectus, blockchain explorers, as well as Telegram channels. Consequently, one avenue for future research is to generalize our findings in a larger sample within a more mature market with a greater variety of determinants, particularly more balanced between conventional and RE STOs for the analysis of underpricing. Besides, we can only consider the returns to investors resulting from the changes in the value of the token and cannot observe and include interest and dividend payments.

Most STOs use the Ethereum blockchain, which merged to the proof-of-stake consensus mechanism in September 2022, silencing criticism of high energy consumption and setting the stage for greater scalability. Hence, this progression will contribute to the future development of the security token industry on a technological and cost level. In many jurisdictions, the record must still be paper-based or stored in a central government database (Lambert et al., 2022). It is necessary for regulators to enact legislation simplifying these processes. Since blockchain technology does not stop at national borders, legislation should ideally be implemented on a large scale, thus ensuring legal certainty for investors.

Appendix

Table A.1: Definition of all variables.

Variable	Description	Source
Pre-STO phase		
Funding amount	Logarithm of the amount of the achieved financing volume in USD	STO research
Funding amount to target	Percentage ratio of the amount of the achieved financing volume to the funding target (Hardcap)	STO research
Transferability	The variable indicates whether a company announces prior the STO that the issued security token will be transferable by the investor (=1), 0 otherwise.	STO research
Equity token	The variable indicates whether the token represents a share in equity (=1), 0 otherwise.	STO research
Fund token	The variable indicates whether the token represents a share in an investment fund (=1), 0 otherwise.	STO research
Voting rights	The variable indicates whether a voting right for the investor is securitized in the token (=1), 0 otherwise.	STO research
Softcap use	The variable indicates whether a funding threshold must be achieved to be completed (=1), 0 otherwise.	STO research
Hardcap	Logarithm of the pre-defined funding target in USD	STO research
Telegram	The variable indicates whether a company uses Telegram as a communication medium with potential investors as part of its STO (=1), 0 otherwise.	Telegram
Listing	The variable indicates whether the company is listed on a traditional stock exchange (=1), 0 otherwise.	STO research
Age	Logarithm of the difference from the start date of the STO and the date of foundation of the company	Own calculations
Cayman Islands	The variable indicates whether the company has been incorporated in the Cayman Islands (= 1), 0 otherwise.	STO research
Europe	The variable indicates whether the company has been incorporated in Europe (= 1), 0 otherwise.	STO research
UK	The variable indicates whether the company has been incorporated in the UK (= 1), 0 otherwise.	STO research
USA	The variable indicates whether the company has been incorporated in the USA (= 1), 0 otherwise.	STO research
Post-STO Phase		
Underpricing	Raw return between token price in the STO and first price on the secondary market	Own calculations
No. large investors	Absolute numbers of investors with a share of more than 5% of all tokens at token issuance	Ethplorer, Etherscan
Sentiment	30-day return of Ether (ETH) on the first trading day	Coinmarketcap
Public float	Percentage share of public float at token issuance	Ethplorer, Etherscan
Trading volume	Logarithm of the trading volume during the first 24 hours on an exchange platform in USD	Exchange Platforms
DEX	Dummy-variable which equals 1 for a decentralized exchange, 0 for a centralized exchange.	STO research
Funding amount	Logarithm of the total funding amount in USD	STO research
Token price	Logarithm of the token price during the offering in USD	STO research
STO type	Dummy-variable which equals 1 for a ‘conventional’	STO research

Continued on Next Page

Variable	**Description**	**Source**
	STO, 0 for a real estate STO (RE STO).	
Sec notation	Dummy-variable which equals 1 for an STO listed on a CEX or DEX, 0 otherwise.	STO research
BHR	Raw buy-and-hold return	Own calculations
BHAR	Buy-and-hold abnormal return adjusted by a value-weighted market capitalization based benchmark	Own calculations
CS estimator, 1 week	Corwin and Schultz (2012) estimator one week after the start of trading averaged over the last five days	Own calculations
CS estimator, 1 month	Corwin and Schultz (2012) estimator one month after the start of trading averaged over the last five days	Own calculations
Liquidity, 1 week	Liquidity measure based on Amihud (2002) and Amihud et al. (2006) illiquidity, one week after the start of trading and averaged over the last five days	Own calculations
Liquidity, 1 month	Liquidity measure based on Amihud (2002) and Amihud et al. (2006) illiquidity, one month after the start of trading and averaged over the last five days	Own calculations
		End of table

Note: List and definitions of all variables with the corresponding source. The source 'STO research' comprises the comprehensive data collection process for the pre-STO phase on Digital Asset Network and various aggregator websites, company websites, EDGAR database, LinkedIn profiles, (legal) prospectus and white papers, blockchain explorers with the corresponding cross-check and for the post-STO phase the exchange platforms.

Table A.2: Correlation matrix for STO success determinants.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Funding amount	1											
(2) Funding amount to target	0.548	1										
(3) Pre−sale	0.087	−0.171	1									
(4) Transferability	0.196	0.143	0.031	1								
(5) Equity token	−0.001	0.048	−0.092	−0.047	1							
(6) Fund token	0.122	−0.118	0.099	0.161	−0.178	1						
(7) Voting rights	0.136	0.135	0.094	−0.078	0.169	−0.169	1					
(8) Softcap use	−0.394	−0.410	0.111	0.075	−0.013	−0.028	0.030	1				
(9) Hardcap	−0.029	−0.317	0.362	0.064	−0.165	0.287	−0.048	0.109	1			
(10) Telegram	−0.149	−0.268	0.492	0.285	−0.038	0.224	−0.024	0.151	0.356	1		
(11) Listing	−0.018	0.070	0.068	−0.216	0.068	−0.087	−0.116	0.045	−0.123	−0.031	1	
(12) Age	0.100	0.181	−0.064	−0.053	0.069	−0.063	0.108	0.005	−0.221	−0.136	0.194	1

Note: This table reports the Bravais-Pearson correlation coefficients for the STO success determinants analysis for the full sample. All variables are defined in Table A.1.

Table A.3: Detailed Descriptives for STO Underpricing.

	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Panel A: Conventional STOs								
Underpricing	20	0.025	0.207	−0.156	−0.135	0.000	0.020	0.490
No. large investors	20	2.200	1.436	1.000	1.000	2.000	3.000	6.000
Sentiment	21	0.047	0.275	−0.583	−0.095	0.105	0.218	0.473
Funding amount	20	14.843	4.481	0.000	13.773	16.321	17.561	18.713
Token price	20	1.862	1.877	0.010	0.693	0.693	2.635	7.311
Trading volume	21	4.334	4.026	0.000	0.000	4.997	7.315	12.219
Public float	20	0.248	0.266	0.000	0.001	0.150	0.455	0.744
DEX	21	0.238	0.436	0.000	0.000	0.000	0.000	1.000
Panel B: Real Estate STOs								
Underpricing	86	0.009	0.126	−0.156	−0.046	−0.022	0.005	0.490
No. large investors	86	3.349	1.532	1.000	2.000	3.000	4.000	10.000
Sentiment	86	−0.012	0.225	−0.583	−0.113	−0.113	0.033	1.289
Funding amount	86	11.632	1.352	10.856	11.021	11.090	11.293	18.421
Token price	86	3.858	0.726	0.693	3.941	3.972	4.010	5.093
Trading volume	86	2.907	1.853	0.000	1.792	2.596	3.886	10.853
Public float	86	0.441	0.325	0.012	0.058	0.629	0.721	0.862
DEX	86	0.965	0.185	0.000	1.000	1.000	1.000	1.000
Panel C: Sample Heckman selection model								
Sec notation	254	0.416	0.494	0.000	0.000	0.000	1.000	1.000
Funding amount	254	11.596	3.600	0.000	10.995	11.220	13.221	18.713

Note: This table reports the descriptive statistics (number of observations, mean, standard deviation, minimum, 25^{th} percentile, median, 75^{th} percentile, and maximum) for conventional STOs (Panel A), Real Estate STOs (Panel B), and the selection equation of the sample for the Heckman selection model for listed and unlisted STOs (Panel C). All variables are defined in Table A.1 in the Appendix.

Table A.4: Correlation matrix for STO Underpricing.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Underpricing	1							
(2) No. large investors	0.280	1						
(3) Sentiment	0.419	0.216	1					
(4) Public float	−0.079	0.150	0.002	1				
(5) Trading volume	0.375	−0.046	0.195	−0.140	1			
(6) DEX	−0.187	0.168	−0.204	0.250	−0.303	1		
(7) Funding amount	0.368	−0.157	0.234	−0.279	0.325	−0.784	1	
(8) Token price	−0.372	−0.036	−0.184	0.243	−0.157	0.716	−0.676	1

Note: This table reports the Bravais-Pearson correlation coefficients for STO Underpricing for the full sample. All variables are defined in Table A.1.

Calculation of the CS estimator

Calculation of the Corwin and Schultz (2012) estimator based on daily high (H_t) and low (L_t) prices of two consecutive time intervals t and $t+1$

$$CS_{t,t+1} = \frac{2(exp(\alpha) - 1)}{1 + exp(\alpha)}$$

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}}$$

$$\beta = \left[ln\left(\frac{H_t}{L_t}\right)\right]^2 + \left[ln\left(\frac{H_{t+1}}{L_{t+1}}\right)\right]^2$$

$$\gamma = \left[ln\left(\frac{H_{t,t+1}}{L_{t,t+1}}\right)\right]^2$$

Chapter 4

Real Estate Security Token Offerings and the Secondary Market: Driven by Crypto Hype or Fundamentals?

This research project is joint work with Ralf Laschinger (University of Regensburg), Bertram I. Steininger (KTH Royal Institute of Technology, Stockholm) and Gregor Dorfleitner (University of Regensburg).

Abstract Tokens, the digital form of assets, are an innovation that has the potential to disrupt how to transfer and own financial instruments. We hand-collected data on 173 real estate tokens in the USA between 2019 and 2021 and trace back 238,433 blockchain transactions. We find that tokens provide broad real estate ownership to many small investors through digital fractional ownership and low entry barriers, while investors do not yet hold well-diversified real estate token portfolios. We analyze the determinants of the success of security token offerings (STOs), secondary market trading, and daily aggregated capital flows. In addition to some property-specific determinants, we find that crypto-market-specific determinants, such as transaction costs and the related sentiment, are relevant both to the STO and capital flows.

Keywords Digital Asset, Security Token Offering (STO), Real Estate Token, Blockchain, Distributed Ledger Technology (DLT), Decentralized Finance

JEL G24, G32, K22, L26, M13

4.1 Introduction

Innovation and technology have influenced and enhanced financial services and products for a long time. One of the most important technical innovations in this context is the *Distributed Ledger Technology* (DLT), a decentralized transparent and tamper-proof verification system.[1] Thus, the blockchain transfers the traditionally centralized ledger system using a single book to the digital world. This technology enables the creation and exchange of digital assets in the form of tokens. Tokenization refers to digitally adding and representing assets in the blockchain (Benedetti and Rodríguez-Garnica, 2023; Schär, 2021). Tokens can be endowed with value, rights, and obligations, similar to traditional forms of ownership, such as stocks or funds. *Smart contracts*, which self-execute once pre-specified conditions are met (Buterin, 2013), enable the issuance and the transfer of tokens time- and cost-efficient. Consequently, financial intermediaries such as banks, exchanges, clearing houses, and notaries are rendered obsolete.

Utility and security tokens can be used to tokenize various rights and assets. Utility tokens grant consumption rights linked to platform services and are issued through an initial coin offering (ICO). Security tokens represent shares of ownership in corporate equity, commodities, currencies, or real estate, and they are issued through a security token offering (STO). After ICOs suffered from a lack of investor protection and frequent fraudulent activities (Momtaz et al., 2019), security tokens emerged as innovative and more trustworthy investment products (Lambert et al., 2022). Security tokens are classified as conventional securities and thus subject to the corresponding regulatory requirements. They can be traded on secondary markets after the offering, enabling divestment and liquidity. The concept of fractional ownership by digital tokens facilitates the fragmentation of assets into multiple tokens, attracting new investors globally to gain access to previously lumpy and illiquid asset classes with high entry barriers. Tokenization is particularly suitable for assets such as land and properties due to their high costs, indivisibility, involvement of multiple intermediaries, and high regulatory requirements (Baum, 2021). Tokens entail lower transaction times since clearing and settlement occur instantly, and costs for third parties (e.g., a broker or notary) is much lower (Ante and Fiedler, 2020; Lambert et al., 2022; Yermack, 2017). This development opens up new diversification opportunities for investors while significantly reducing costs and illiquidity premia, paving the way toward entirely digitized financial markets.

The financial industry has already developed various solutions for (in-)direct investments in real estate due to the attractive characteristics of real estate in terms of constant cash flows or low correlation to stocks and bonds. Specifically, open and closed-end funds or

[1] In this article, we employ the terms DLT and blockchain synonymously, even though the blockchain represents only one subtype of DLT. For a detailed discussion, see Liu et al. (2020b).

REITs enable retail investors to gain access to this asset class. The increasing adoption of blockchain has led to the emergence of real estate tokens as a new investment vehicle and digital surrogate for direct property ownership (Baum, 2021). A real estate token, like closed-end funds, mostly comprises one property and not a portfolio of properties, such as open-end funds and REITs. In the case of REITs or funds, investors do not own the properties and, unlike tokens, cannot influence the decision to invest in a particular property. A token gives the investor fractional ownership of the property, making it the technically closest form to fractional direct investment to date. In contrast to closed-end funds, token investors can avoid high minimum investment amounts and administrative costs.

The literature on real estate tokens is to date mainly of a theoretical nature regarding the general procedure (Gupta et al., 2020; Liu et al., 2020a; Markheim and Berentsen, 2021), financial application (Baum, 2021; Markheim and Berentsen, 2021), legal (Konashevych, 2020), and technical aspects (Gupta et al., 2020). Markheim and Berentsen (2021) present descriptive data based on a small sample of real estate tokens, where they point, despite the many theoretical advantages of tokens, towards challenges, such as regulatory uncertainties and relatively long transaction times. Swinkels (2023) examines the liquidity and ownership of real estate tokens using the same data source as our study, albeit with an earlier end date, and considers 58 tokens. His findings suggest that a tokenized property has, in the mean, 254 owners, with ownership changes occurring annually on average. In addition, he concludes that investors are interested in the exposure to the residential house price index, as token prices are linked to housing prices. Our study starts one step earlier and differentiates between the determinants of STOs on the transaction level and daily capital flows on the macro level.

We hand-collected data on 173 real estate tokens with their property and financial characteristics in the USA between 2019 and 2021. Moreover, we examine the related 238,433 blockchain transactions to analyze investor behavior. We have enriched this database with crypto market-specific characteristics and macroeconomic indicators. In this regard, our main findings are threefold.

First, we are among the first to trace back the underlying blockchain transactions in an empirical analysis to derive insights into investor behavior. Our analysis shows that investors hold an average of ten different tokens and an investment amount of 4,030 USD, which does not represent a well-diversified real estate token portfolio. Tokenization provides broad access to real estate ownership for many small investors as property ownership is not concentrated on a few large investors. Most investors acquire tokens during STOs, while secondary market trading plays a minor role. Second, we investigate the determinants of STO success, defined as the number of days until all tokens are sold and the mean funding amount per day. For the latter and primary success variable of interest in this study, we find that some property-specific fundamentals and the crypto market-related transaction costs explain most of the success of the STO. Third, we switch from the individual STO to the macro-level view of aggregated daily capital flows per property to account for the specific crypto market over time. We observe that real estate token investors similarly consider the crypto market sentiment and transaction costs when purchasing tokens. In contrast, only transaction costs directly reducing the return on investment are relevant when selling. Additionally, macroeconomic factors have a minor role in capital flows.

Our study contributes to several streams of literature. First, we add to the literature on blockchain technology and the economics of digital assets. The first wave of academic literature in this sub-stream focused on ICOs as an innovative form of crowdfunding, bearing the advantage that the blockchain tokens enable secondary market trading (Lee et al., 2022). Empirical studies on ICOs examine success determinants (Fisch, 2019; Howell et al., 2020), investor characteristics and motives (Fisch et al., 2021; Fahlenbrach and Frattaroli, 2021), white papers (Florysiak and Schandlbauer, 2022; Thewissen et al., 2022), and post-ICO performance (Benedetti and Kostovetsky, 2021; Fisch and Momtaz, 2020; Lyandres et al., 2022). Momtaz (2023) emphasizes that the reasons security tokens are driving digitization in finance are interoperability, fractional ownership, instantaneous settlement, and market liquidity. Gan et al. (2021) find that STOs, in contrast to ICOs, entail lower agency costs, lower token turnover, lower cash diversion, and raise higher amounts of funds and firm profits. The existing empirical literature on STOs primarily examines success determinants during the funding process, focusing on the issuer and offering characteristics (Lambert et al., 2022; Ante and Fiedler, 2020).

Second, we contribute to the literature on real estate investments. The real estate sector is a major sector for study in its own right in the literature on crowdfunding (Jiang et al., 2020; Schweizer and Zhou, 2017; Shahrokhi and Parhizgari, 2020). Fisch et al. (2022) compare ICOs and, among others, REITs to analyze whether gender, ethnicity, and geography influence the decision for an ICO. While the authors point out that real estate is a highly relevant use case for blockchain-based financing, they do not directly examine real estate STOs. In a sample of 1,125 ICOs for external firm financing, Howell et al. (2020) find a positive relationship between ICO success, measured by employment, and the operating sector of tokenizing real assets. They attribute this result to the underlying concept of security tokens but do not deepen the analysis further on this aspect. STOs of real estate projects need to be studied separately to simultaneously consider the underlying asset class and the specific crypto market environment.

Third, we complement the literature on portfolio construction and diversification. Diversification is a fundamental concept in portfolio theory (Markowitz, 1952). Goetzmann and Kumar (2008) document that 60,000 individual US investors hold under-diversified equity portfolios, leading to high idiosyncratic risk and, consequently, a welfare loss. The small investment amount resulting from fractional ownership of digital tokens theoretically makes diversification easier. Therefore, we aim to verify whether real estate tokens live up to their promise of portfolio diversification.

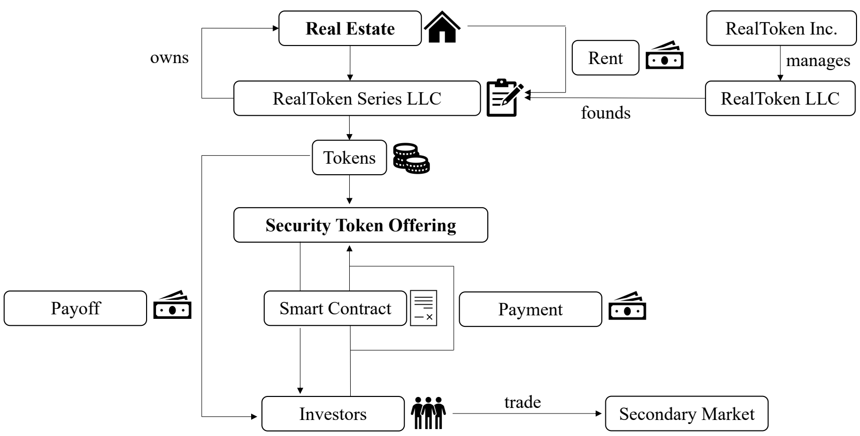
The remainder of this paper is organized as follows. In Section 4.2, we present the real estate tokenization process and derive our hypotheses. We describe our data and method in Section 4.3. The main analyses and discussion of our empirical results are presented in Section 4.4, followed by further analyses and robustness checks in Section 4.5. In Section 4.6, we conclude our study.

4.2 Conceptual framework and derivation of hypotheses

4.2.1 Real estate tokenization

Our dataset comprises real estate tokens issued by the platform RealToken (RealT), an active issuer and platform for real estate tokens in the USA. Based on the *Howey test*, digital assets are investment contracts and, therefore, considered securities. Consequently, real estate tokens must be registered with the Securities and Exchange Commission and are subject to laws and regulations protecting investors. RealT offers the tokens in unregistered securities offerings, or private placements, under Regulation D 506(c) (US-accredited investors) and Regulation S (non-US investors) of the Securities Act. We illustrate the process of real estate tokenization and STOs in the case of RealT in Figure 4.1 and describe the process below.[2]

Figure 4.1: Process Map



Note: This figure illustrates the process of real estate tokenization and STOs in the case of the platform RealT.

[2]For a description of the ICO or STO process, see Momtaz (2020) and Lambert et al. (2022).

RealToken LLC creates a RealToken Series LLC for each property since properties cannot be directly digitalized. This LLC acts as a special purpose vehicle (SPV) and holds the property deed.[3] These SPVs stand solely and legally on their own and are, in the next step, tokenized using the technical standard of the Ethereum ERC-20 token. The properties are primarily rented residential buildings. Property management is outsourced to local professionals. Investors can purchase the tokens during the STO. After successful payment and signing the offering memorandum digitally, they automatically receive the tokens in their wallets employing a smart contract. On the Ethereum blockchain, computing power is required to perform operations successfully, and users have to additionally pay a so-called *gas fee*. The tokens give the investor a deed in the respective tokenized RealToken Series LLC. After operating costs, insurance, and real estate taxes, the net rent is submitted weekly to the RealToken rent contract linked to the property and automatically issued to the token holders' wallets. The value of a token is specified by the assessed property value after a maintenance and repair reserve divided by the total number of tokens issued. RealT charges a fee of 10%, for which investors, in exchange, receive governance tokens from RealT itself. Afterward, the security tokens can be either returned to RealT or traded on decentralized exchanges (DEX) as a means of decentralized finance.[4] The properties are re-valued annually, resulting in the depreciation or appreciation of the tokens. After the rapid increase in transaction costs in combination with longer execution times on the Ethereum blockchain at the beginning of 2021, RealT decided to alternatively enable transactions on the Gnosis blockchain.[5] In particular, for the relatively low weekly rent payments, using Gnosis and avoiding high transaction costs on the Ethereum blockchain is favorable. After elucidating the mechanics of real estate tokenization, the following hypotheses are derived from the academic literature.

4.2.2 Derivation of hypotheses

We first tackle the impact of different property-specific factors on the perceived quality, risk, and expected cash flow, which can be related to the success of an offering. From a theoretical perspective, property type and location are the major property-specific characteristics that influence value. These factors are empirically confirmed by various studies (see, e.g., Cronqvist et al., 2001; Pai and Geltner, 2007; Ro and Ziobrowski, 2012; Hartzell et al., 2014). Real estate is naturally immobile, which means that the location determines its value to a large extent. Therefore, a purchaser acquires both the building and the site at

[3] A form to digitize ownership is non-fungible tokens (NFTs) or with the help of Decentralized Autonomous Organizations (DAOs). However, these are only theoretical concepts not often applied to the real estate market and, consequently, lie beyond the scope of this paper.

[4] For a detailed discussion, see Aspris et al. (2021).

[5] Gnosis (formerly xDai) blockchain is a second-layer protocol to create, trade, and hold digital assets on Ethereum.

the same time (Kiel and Zabel, 2008). The options for determining the location's quality are manifold: political or historical zones, indirect factors, such as the school quality of the district, or the distance to important places, such as the central business district. These indicator variables mostly imply indirect influences on house values since investors consider specific locations or location characteristics more or less favorable. In particular, the low minimum investment amount for tokens enables investors to diversify their portfolios more broadly, especially regarding location. This makes the location an important factor for the attractiveness of the STO for an investor and could, consequently, influence the success of a real estate STO.

The size of the property measured by its value determines the rent and return, similar to the way the size factor determines the return on the stock market (Fama and French, 1993). Geltner et al. (2014) report that size is a suitable factor for explaining the return variation of real estate on a large scale. Pai and Geltner (2007) use the market value as a size factor and find the opposite impact compared to the stock market – larger properties have a higher expected return premium. Esrig et al. (2011) state that large properties outperform other properties on an absolute and risk-adjusted basis for different property types. Sirmans et al. (2005) conduct a review of around 125 studies using hedonic modeling to estimate house prices and report that lot size had a positive effect in the vast majority of observations. Therefore, we expect that the size has a positive relationship with the success of the STO.

If the quality of the property is not specified, its age can be used as a proxy for it. A lower quality induces higher uncertainty for maintenance and repair costs and, thus, higher risk for the buyer (Bourassa et al., 2009). Since investors try to avoid this kind of risk, older properties may be less attractive to investors. This argumentation is supported by Sirmans et al. (2005), who find in their review that the influence of age on house prices was almost entirely negative.

The major risk regarding the expected cash flow is a rent default. This risk can be reduced by splitting the rent between several different tenants. Therefore, single-tenant buildings limit the diversification possibilities of potential investors in contrast to multi-tenant properties. The limited diversification options make single-tenant properties, in contrast to multi-tenant properties, less attractive, which may result in a less successful funding process. Opposed to that Ling and Archer (2021) find that single-family properties have a lower risk than multi-family homes because single-family homes are typically located in desirable suburban areas with steady demand. Based on the importance of both effects – lower default risk for multi-tenant buildings vs. location – an exact expectation cannot be formulated, and the issue has to be settled empirically.

In the USA, low-income households can receive rental housing assistance via Section 8 of the United States Housing Act of 1937 (42 U.S.C. §1437 et seq.). This program helps them in finding a decent and affordable place to live. The state pays the rent directly to the landlord, which significantly reduces the risk of payment issues or default. The Section 8 program guarantees token purchasers a stable and predictable rent payment. Consequently, investing in such properties bears a lower risk of rent default. Investors may find properties with a greater percentage of rental assistance from the Section 8 program to be more attractive. As such, our Hypothesis 1 reads:

> **Hypothesis 1:** *The quality of a location, the size of a property, and a higher portion of rental assistance through Section 8 are positively related to the success of an STO, while age is negatively related.*

In addition to the property and financial characteristics, we also consider campaign features commonly known from the literature on crowdfunding (CF) (Belleflamme et al., 2014). In the context of CF, it is decisive for the funding success of a campaign to be able to signal the quality of a project to potential investors (Ahlers et al., 2015). Conventional CF campaigns often have a short or missing track record or lack a market-ready product. Therefore, investors need to base their decision on other information, such as the description in text and pictures on the platforms. This information allows companies to reduce information asymmetries and signal project quality (Diamond, 1984). Apart from the text, pictures assist in visualization and enable an evaluation of the property's location and actual condition. Previous CF studies identified a detailed project description to overcome information asymmetries and increase campaign success (De Crescenzo et al., 2020; Gao et al., 2023). This effect has also been investigated in the literature on real estate for its impact on home prices and home-buyer attention in a similar vein (Luchtenberg et al., 2019; Nowak and Smith, 2016; Seiler et al., 2012). The more detailed and larger the number of pictures, the more realistic and accurate the presentation of the potential investment is for an investor. High-quality projects are incentivized to deploy detailed project descriptions, whereas low-quality projects tend to be vaguer in their disclosures. Therefore, we assume that a detailed project description is a positive quality signal for an investor, which prompts an investment and can increase the success of an offering.

> **Hypothesis 2:** *A detailed project description is positively related to the success of an STO.*

The investment decision process, akin to other markets, is potentially driven by the market-specific environment and investor or market sentiment. Investors follow investment recommendations and central strategies, and retail investors mostly exhibit herding behavior, often caused by market sentiment. Herding behavior has been studied in the traditional stock market (Chang et al., 2000; Chiang and Zheng, 2010; Litimi et al., 2016) and in the cryptocurrency market (Ajaz and Kumar, 2018; Bouri et al., 2019). Investors, particularly non-rational investors like many crypto investors, are potentially subject to herding behavior. Investor sentiment can be particularly pronounced in the market for tokens (Drobetz et al., 2019), as this seems to be in such highly subjective asset classes (Baker and Wurgler, 2006). From an investor perspective, we assume, similarly to Ante and Fiedler (2020), that in the market for STOs, a house money effect exists, meaning that investors take higher risks after prior gains (Thaler and Johnson, 1990), especially during periods of positive market sentiment. Since issuers anticipate this irrational investor behavior, they will await the right time on the market to place the offers. For example, Drobetz et al. (2019) show that companies seeking funding via ICOs avoid phases of general negative market sentiment for their exchange listing, which results in short-term negative returns of the tokens. Token platform operators can time the publication of a project to periods of positive market sentiment. Thus we expect a positive link between market sentiment and the success and daily capital inflows as token purchases and a negative link with daily capital outflows as token sales.

With regard to the specific market environment for blockchain-based tokens, a cost effect that runs counter to the market sentiment must also be taken into account. Apart from the administrative fees directly imposed by the token issuer, specific transaction costs called *gas fees* are additional costs associated with a token investment that need to be considered and paid by the investor. Since gas is needed to perform operations and space is limited on a block, the resulting transaction costs may vary due to fluctuations in supply and demand on the network.[6] Gas fees rise when demand increases, and vice versa; hence, they signify crypto popularity. Additionally, users can pay an extra fee to increase the likelihood of their transaction being included in the next block when demand is high. Gas fees can be observed and predicted easily for investors on corresponding websites opening up the possibility to time the investment and avoid high transaction costs. Momtaz et al. (2022) provide the first empirical evidence of tokens on the Ethereum blockchain, including stablecoins, startup tokens, and lottery tokens. The authors find that investors reduce their trading activity when transaction costs are high. In conclusion, we expect that crypto market transaction costs are negatively related to the success of an STO and capital inflows and outflows because investors seek to circumvent high transaction costs. The decision of an investor to make a real estate token investment can therefore be based on two opposing effects as indicators of crypto popularity, which is why an empirical investigation is required.

> **Hypothesis 3a:** *Crypto market sentiment is positively related to capital inflows, while it is negatively related to capital outflows.*

[6] By definition, 'gas fee' and 'transaction fee' are not synonyms, as the actual total cost per transaction is the multiplication of gas used and a base gas fee. For more detailed information on the mechanism and calculation of gas fees, see Ethereum.org (2022).

Hypothesis 3b: *Crypto-market related transaction costs are negatively related to the success of an STO as well as capital inflows and outflows.*

4.3 Data and method

4.3.1 Data sources

We collect the US real estate token data directly from the RealToken platform, resulting in 173 financed projects as of December 31, 2021. The data comprises information at the property level and its financial characteristics. The blockchain transaction data comes from two blockchain explorer and analytic platforms, namely *Blockscout* and *Etherscan*, which was also used by Lyandres et al. (2022). We rely on these two sources for the transaction data as RealT has enabled transactions on the Gnosis blockchain since the beginning of 2021.

4.3.2 Method: blockchain transaction analysis

The blockchain is a digital ledger in which one entry corresponds to one transaction. We derive all blockchain transactions related to the real estate tokens in our sample until the end of our observation period in December 2021. The structure of a blockchain transaction comprises the respective token, a unique transaction hash (transaction ID), a time stamp, the number of tokens, and the sending (from) and receiving addresses (to). We trace back investors through their unique and pseudonymous *wallet* address, which is comparable to the account number in the traditional banking sector. Even if an investor can have several wallets and, thus, more than one unique wallet address, we assume that most investors have only one wallet.[7] The switch of the blockchain from Ethereum to Gnosis is no issue regarding the unique wallet address, as Gnosis is built upon Ethereum and, therefore, the wallet addresses remain the same. Due to the focus of our study, we do not consider other investments by investors in their wallets besides real estate tokens. We can clearly distinguish transactions from the STO from secondary market transactions by identifying the emitting wallet address of the platform operator from which tokens are transferred to investors for each property. Consequently, the remaining transactions from non-emitting wallets are secondary market buy-or-sell transactions.

Based on the transaction data, we derive several variables that shed light on both investors and their investment strategies concerning tokenized properties. To this end, we analyze

[7] This assumption can be justified for several reasons. On the RealT platform, a user can only deposit one wallet at a time. Swinkels (2023) has submitted a request to the platform operator confirming the assumption. From an academic point of view, Fahlenbrach and Frattaroli (2021) have conducted tests in an ICO sample and found similar results.

two distinct perspectives: the wallet-investor and the token-property perspective. In the wallet-investor perspective, the variable *Properties per Investor* accounts for the number of properties an investor has invested in. This variable addresses the extent to which investors diversify their real estate token portfolio. Further, we convert the number of tokens observed in the transactions into a more easily interpretable and meaningful dollar amount, using the price of the tokens from the STO and calculate the *Holdings per Investor as of Dec 2021* in dollars. To measure the time dimension of the investments and thus the willingness to speculate on the side of the investors, we analyze the *Holding Period all Investors as of Dec 2021* in days. From the token-property perspective, we consider the concentration of ownership with the Herfindahl-Hirschman index (Herfindahl, 1950; Hirschman, 1964). We calculate the Herfindahl-Hirschman index as

$$HHI = \sum_{i=1}^{N} s_i^2 \tag{4.1}$$

in which s is the percentage of ownership of an investor i, and N constitutes the total number of investors on the property level. The index ranges between $1/N$ and 1. The latter implies that complete ownership is concentrated on a single investor. To account for variations in the HHI caused by a different number of investors in the properties and to facilitate direct comparison between properties, we consider the normalized Herfindahl-Hirschman index as

$$HHI^* = \frac{HHI - 1/N}{1 - 1/N}. \tag{4.2}$$

This measure varies between 0, which corresponds to equal ownership of all investors, and 1, which corresponds to a single investor with full ownership. The variable *Investors per Property* measures the number of unique wallets invested in a specific property.

In addition, we examine investors' trading activities on both the buy and sell sides. With the variable *STO Buy*, we measure the absolute dollar amount of purchases during the STO. Figure A.1 in the Appendix illustrates the calculation scheme of the *Secondary Market Buy* and the *Secondary Market Sell* side. We implemented a daily balance calculation to summarize the transactions per day to determine the daily dollar holdings per wallet. This approach entails evaluating the changes in wallet balances over time, where an increase in the balance indicates a buy transaction, and a decrease in the balance represents a sell transaction. We use this method because the dollar volume per wallet gives more insight than the volume per individual transaction. Therefore, *Secondary Market Buy* depicts how large the purchasing investment amounts are in the secondary market. The variable *Secondary Market Buy/Existing Exposure* indicates the percentage of purchases on the secondary market compared to the existing investment. On the sell side, we analyze with the variable *Secondary Market Sell* the dollar amount investors sell on the secondary market. The variable *Secondary Market Sell/Existing Exposure* puts this in relation to the existing investment. Lastly, the variable *Holding Period Sellers* measures how many days investors who sell their tokens have previously held them. The latter two variables provide insights into whether investors are interested in regular cash flows from the rent payments or the changes in the token's value itself.

4.3.3 Method: multivariate analysis STO success determinants

In the first multivariate analysis, we test for determinants of the success of real estate STOs. We operationalize the funding time and speed as our measures of success. The funding time measures the number of days until 95% of the tokens have been transferred to the investors' wallets since RealT retains tokens to ensure liquidity in the secondary market, based on the blockchain transaction data.[8] Therefore, it is a proxy for the pure time dimension of success. We consider a project more successful if it takes less time to secure funding. We sub-categorize the funding time into the *Funding Time until Success* for the sub-sample of successfully funded projects transferred to the investors' wallets. As the second sub-category of funding time, we simultaneously examine successful and unsuccessful projects regarding the *Funding Time until Dec 2021* to obtain a sample free of survivorship bias. We estimate the parametric accelerated failure-time (AFT) survival model to account for unsuccessful projects correctly and because the proportional hazards assumption is violated for the semi-parametric Cox model. We apply the lognormal and log-logistic distributions since both present the most appropriate statistical fit for the distribution of our dependent variable. The AFT model is an alternative to modeling survival times often used in crowdfunding (Jiang et al., 2020; Felipe et al., 2022).

The funding time may be positively related to higher amounts of *Total Investment*. Therefore, we alternatively consider the measure speed. It is the fraction of 95% of the *Total Investment* to the funding time. Thus speed measures the mean investment amount funded per day.[9] Successful projects have a higher speed, corresponding to a higher daily funding amount. Analogously to the analysis of the funding time, we sub-categorize speed in the first specification with the corresponding *Funding Time until Success* into the dependent variable *Speed until Success* for successful projects. In the second model specification, we examine all projects as of December 2021 with *Speed until Dec 2021*. For projects that have not been successfully funded until the end of our observation period and are on the market longer than the mean time of *Funding Time until Success*, we equate *Speed until Dec 2021* to 0 to proxy a low speed and prevent distortions from unsuccessful projects with

[8] In Subsection 4.5.1, we vary and verify the 95% assumption for an STO in order for it to be considered successful.

[9] This definition is analogous to the average velocity in physics, based on the investment amount instead of distance.

a large *Total Investment*. For projects that have not been successfully funded until the end of the observation period and are on the market shorter than the mean time of *Funding Time until Success*, we use the actual amount of money raised instead of *Total Investment*.

In the baseline regression, we include the financial, property, and campaign variables which we expand in the second specification with crypto market-specific characteristics. We use robust standard errors that are one-way-clustered in all regressions and quarter-year dummy variables. The financial characteristics of the property include *Rent per Token p.a.* for the annual rent a token holder receives per token. The variables *Expected Yield* and *Total Investment* are data publicly available before funding. These variables are determined by the property characteristics and thus can be indirectly influenced by the token issuer. The financial ratio *Expected Yield* is given by the ratio of the net rent to the token price. *Total Investment* refers to the amount of money needed to secure successful funding. This variable is commonly used in the CF (Block et al., 2018; Mollick, 2014), ICO (Adhami et al., 2018; Fisch, 2019), and STO literature (Ante and Fiedler, 2020; Lambert et al., 2022) to determine project success and represents the funding amount actually collected. However, due to the technical procedure on the blockchain, the *Total Investment* in our context is always entirely issued as part of tokenization but not necessarily fully transferred to investors. At the same time, the issuer keeps the remaining tokens. Therefore, we do not apply this variable as a measure of success.

The property characteristics comprise the variables *Age*, *Lot Size*, *Section 8* as the percentage of the share of financially supported housing within one property, and the type of use with the dummy variable *Single Family* if one family is the only tenant. For a suitable location variable, we rely on the dummy variable *Detroit* and the metric variable *Distance DTWN* to account for location quality since these variables are easily accessible and straightforward to understand for a retail investor. Similar to Swinkels (2023), we assume that rental properties outside of Detroit are more attractive for investors for diversification reasons, as the majority are located in Detroit. In addition, we also measure the distance to downtown in miles with the variable *Distance DTWN* to incorporate the micro-effects of the location. The campaign characteristics related to the literature on crowdfunding include the number of pictures with the variable *#Pictures* and the length of the descriptive text with *#Characters* for the particular property project.

For market-specific variables, we include for the crypto environment the variable *Gas Fees* for transaction costs on the Ethereum blockchain, converted to USD. Additionally, we include the S&P Case-Shiller Home Price Index with the variable *Housing Market* for the respective regions corresponding to the particular cities where the properties in our sample are located (Detroit, Chicago, Cleveland, New York, and Florida), lagged for one month. Since investors participate in the value depreciation or appreciation of the property with the value of their token, they care about the growth potential of the real estate market.

They may be more willing to purchase a token if the regional real estate market grows. All variables are defined in Table A.1 in the Appendix.

4.3.4 Method: multivariate analysis funding determinants

With the multivariate analysis of STO success determinants, we analyze the STO at that specific point in time. However, when considering the crypto market over time, we must detach from mostly time-invariant STO characteristics and move on to the macro-level view of real estate token market activity. Hence, we can additionally account for daily fluctuations, notably for short-term particularities and shocks. In concrete terms, this shifts our models from the STO perspective to a daily view of capital inflows and outflows over time. To account for unobserved effects regarding individual characteristics and time, we employ a two-way fixed effects panel regression to analyze the determinants of daily inflows and outflows per property.

The dependent variables, daily *Inflow* and *Outflow* per property, are calculated based on the blockchain transaction data. *Inflow* indicates how much money investors spent during the STO or on the secondary market per property on a given day. *Outflow* measures which amount of money the investors sold from a property on the secondary market on a given day.[10] The *Inflows* and *Outflows* in the market for real estate tokens may be influenced by determinants and shocks both in the crypto market and the macroeconomy. Therefore, to account for the peculiarities of the crypto market, we consider from the sentiment perspective the five-day cumulative return of the native token of the Ethereum blockchain, Ether (ETH), with the variable *ETH Price* denominated in USD. The market capitalization of ETH is the second largest after Bitcoin on the cryptocurrency market as of December 31, 2021, and Ethereum is the primary platform for security tokens. Since the cryptocurrency market is still in its infancy and the general conditions are changing, it is characterized by high volatility. To incorporate short-term shocks in the crypto market, we include the dummy variables *ETH Shock* and *Gas Shock*. *ETH Shock* equals one if the cumulative return of five days prior to the observation decreased by more than 5% and *Gas Shock* which equals one if the cumulative return of *Gas Fees* increased by more than 5% in five days. For the macroeconomic environment, we include the *One-month Treasury*, *Ten-year Treasury*, and the *Aruoba-Diebold-Scotti* Business Conditions Index (*ADS Index*) of Aruoba et al. (2009). According to the Federal Reserve Bank of Philadelphia, the *ADS Index* covers seasonally adjusted macroeconomic indicators, including, among others, initial jobless claims (weekly), payroll employment (monthly), industrial production (monthly),

[10] *Inflows* are the aggregated *STO Buy* and *Secondary Market Buy* and *Outflow* is *Secondary Market Sell* per project, reported in Table 4.2. The difference in the number of observations is due to the fact that there were no sales on some days. The difference in the mean is caused by *STO Buy* transactions and represents the capital that investors actively hold in tokens.

and real GDP (quarterly). The index offers the advantage that, unlike, e.g., GDP or the unemployment rate, the data is provided daily, corresponding to the daily frequency of our dependent variables. Due to its high frequency, the index is increasingly used in academic research (see, e.g., Caporin et al., 2022; Da et al., 2014).

4.3.5 Descriptive statistics

Table 4.1: Descriptive Statistics

	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Panel A: Variables for STO success determinants								
Dependent variables								
Funding Success	173	0.72	0.45	0	0	1	1	1
Funding Time until Success	125	48.72	49.53	2.63	9.87	26.92	82.29	226.70
Funding Time until Dec 2021	173	73.01	67.03	2.63	11.91	56.53	121.00	323.00
Speed until Success	125	10.55	20.38	0.27	0.95	4.19	9.22	128.48
Speed until Dec 2021	173	8.30	18.45	0.00	0.27	1.78	8.64	128.48
Explanatory variables								
Rent per Token p.a.	173	5.98	1.59	3.96	5.53	5.81	6.08	21.82
Total Investment	173	168.02	205.54	48.08	60.58	66.50	144.45	985.91
Expected Yield	173	0.11	0.01	0.07	0.11	0.11	0.12	0.13
Age	171	85.02	18.48	2	74	84	94.5	134
Lot Size	166	5,338.20	2,951.67	871	3,920	4,792	5,644.5	29,620
Section 8	173	0.18	0.37	0.00	0.00	0.00	0.00	1.00
Single Family	173	0.64	0.48	0	0	1	1	1
Distance DTWN	173	4.70	1.73	1.08	3.61	4.51	5.40	9.63
Detroit	173	0.80	0.40	0	1	1	1	1
#Pictures	173	4.34	4.77	1	2	3	5	35
#Characters	172	205.65	305.82	0	0	0	364.2	1,654
Gas Fees	173	6.68	4.53	1.11	1.78	6.79	9.42	16.85
Housing Market	173	150.67	24.35	127.56	139.63	148.45	155.38	343.64
Panel B: Variables for funding determinants								
Dependent variables								
Inflow	26,940	1,189.39	11,201.43	0.00	5.00	16.01	117.98	493,278.80
Outflow	26,016	218.44	1,484.38	0.00	4.87	13.80	65.50	71,819.98
Explanatory variables								
Gas Fees	654	4.37	4.23	0.76	1.41	1.78	8.11	18.00
ETH Price	654	1,266.96	1,392.49	110.61	202.23	387.98	2,232.96	4,812.09
Gas Shock	654	0.38	0.49	0	0	0	1	1
ETH Shock	654	0.30	0.46	0	0	0	1	1
One-month Treasury	627	0.53	0.77	0.00	0.05	0.09	1.52	2.26
Ten-year Treasury	627	1.29	0.44	0.52	0.84	1.43	1.63	2.13
ADS Index	654	−0.47	5.64	−26.33	−0.31	0.18	0.86	8.99

Note: This table reports the descriptive statistics (number of observations, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum) for the full sample. For the analysis of STO success determinants, the number of observations of 125 of *Funding Time until Success* and *Speed until Success* refers to the successful projects in the sample; the remaining variables represent the entire sample of 173 observations. For the analysis of the funding determinants, the number of observations differs between *One-month Treasury*, *Ten-year Treasury*, and the remaining explanatory variables, as these data are not provided on bank holidays. All variables are defined in Table A.1 in the Appendix.

The descriptive statistics for analyzing success determinants are displayed in Table 4.1 in Panel A. In our total sample of 173 real estate STOs, 72% were successful, which indicates that 95% of the tokens were transferred to investors. The sub-sample of successful STOs has a mean *Funding Time until Success* of 48.72 days and a median value of 26.92 days. In contrast, the *Funding Time until Dec 2021* for the entire sample is correspondingly longer, with 73.01 days in the mean. The minimum of 2.63 indicates that some very attractive projects sell off quickly. The money-oriented variable *Speed until Success* has a mean of 10,550 USD/day for successful projects and a median of 4,190 USD/day. When considering successful and unsuccessful projects regarding the *Speed until Dec 2021*, the mean of 8,300 USD/day is subsequently lower. The minimum *Speed until Dec 2021* of 0 represents projects not fully funded within the mean of *Funding Time until Success* of 48.72 days.

For the *Expected Yield*, the mean is at 11%. The mean property value measured by the highly skewed *Total Investment* at 168,020 with a median of 66,500 shows that most properties have a relatively low value. Among the housing characteristics, we observe that 80% of the properties are located in *Detroit* and 64% are *Single Family*. The campaign variables show that the offers, on average, are illustrated with four pictures and described in 205.65 characters. We do not consider the variable *#Characters* further in our multivariate analysis since the median value is zero because the platform did not provide any descriptive text at the beginning. The *Gas Fees* at the day of the STO range from a minimum of 1.11 to a maximum of 16.85, with a mean of 6.68, highlighting that blockchain-related transaction costs fluctuate and can be of crucial interest to token investors.

Panel B presents the descriptive statistics for the analysis of funding determinants. The unbalanced panel data set consists of 26,940 daily *Inflow* and 26,016 daily *Outflow* observations per property per day over our observation period of about two and a half years as of December 2021. On average, *Inflows* have a mean of 1,189.39, highly distorted by the maximum of 493,278.90 from an expensive and quickly sold property. The daily *Outflows* per property amount to a mean of 218.44. The medians of daily *Inflows* and *Outflows* are in a similar magnitude range at 16.01 and 13.80. The daily *Gas Fees* range between a minimum of 0.76 and a maximum of 18.00 throughout the observation period. The mean of *ETH Price* is 1,266.96 with a median of 387.98. The latter two variables illustrate the high volatility of the crypto market, which is why an additional examination of short-term shocks is required. A *Gas shock* is present in 38% and a *ETH Shock* in 30% of the daily observations. Table A.2 in the Appendix displays the Bravais-Pearson correlation coefficients for all of the variables we consider in the analysis of STO determinants. The correlation coefficients between the explanatory variables are moderate and provide initial evidence for our hypotheses.

4.4 Main analyses

4.4.1 Analysis of blockchain transaction

Table 4.2: Blockchain Transaction Analysis.

	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Panel A: Wallet-Investor perspective								
Properties per Investor	6,806	10.2	20.7	1	1	3	9	171
Holdings per Investor as of Dec 2021	6,544	4,029.35	32,319.99	0.00	57.96	259.45	1,398.34	1,439,474.00
Holding Period all Investors as of Dec 2021	165,161	244.51	160.59	0	133	221	286	850
Panel B: Token-Property perspective								
HHI* STO	173	0.03	0.06	0.01	0.01	0.02	0.04	0.68
HHI* as of Dec 2021	172	0.03	0.04	0.01	0.01	0.02	0.04	0.28
Investors per Property	173	401.2	201.2	31	258	328	501	1,173
Panel C: Buy side								
STO Buy	87,048	317.82	2,467.28	0.00	35.98	57.96	162.60	155,010.00
Secondary Market Buy	35,351	88.70	721.13	0.00	2.92	6.72	25.43	58,462.74
Secondary Market Buy/Existing Exposure	35,351	0.38	11.67	0.00	0.01	0.03	0.10	2,104.88
Panel D: Sell side								
Secondary Market Sell	31,697	99.97	802.28	0.00	3.00	7.65	25.69	58,462.74
Secondary Market Sell/Existing Exposure	31,697	0.09	0.16	0.00	0.01	0.02	0.07	1.00
Holding Period Sellers	31,638	105.09	86.06	1	36	86	155	701

Note: This table reports the descriptive statistics (number of observations, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum) for the wallet-investor perspective (Panel A), token-property perspective (Panel B), as well as the buy side (Panel C) and sell side (Panel D). The sample includes 238,433 blockchain transactions from 2019 to 2021. Figure A.1 in the Appendix illustrates the calculation scheme of the buy and the sell side. All variables are defined in Table A.1 in the Appendix.

Based on 238,433 blockchain transactions related to all real estate tokens in our sample, we identify 6,806 unique wallets representing the corresponding number of real estate token investors. The different number of observations per variable is due to different transactions and filtering methods, both of which serve to derive the respective variable of interest. From the wallet-investor perspective in Table 4.2 in Panel A, we document that one single investor invests in 10.2 properties on average. However, at least 25% of all investors have invested in only one property. One reason for this observation could be the novelty and peculiarity of real estate tokens. The respective investors do not yet hold a diversified real estate token portfolio. This result is in line with a previous study of ICO investors which finds that the main reason for a token investment is technological motives, followed by financial reasoning (Fisch et al., 2021). The maximum of 171 distinct properties out of 173 exemplifies that there are also investors who have invested in almost every property and have well-diversified tokenized real estate portfolios.[11] After converting the number of tokens into dollar holding amounts, we find that the mean of *Holdings per Investors as of Dec 2021* is 4,029.35 USD, and the median is 259.45 USD. The mean of *Holding Period*

[11] Due to the pseudonymity of wallets on the blockchain and the fact that we can only trace back the issuing wallets of RealT, we cannot completely rule out the possibility that our maxima are influenced by other wallets used for handlings and shifts by the token issuer.

all Investors as of Dec 2021 is 244.51 days with a maximum of 850 days, indicating that investors of the first STO are still holding the tokens.

If we switch to the token-property perspective in Panel B, we see a high dispersion and less concentration of ownership based on the mean of the normalized *HHI** of 0.03 both after the STO and as of December 2021. This result indicates that not only a few investors hold the majority of tokens, but that tokenization, in practice, provides broad access to real estate ownership for many small investors. This result aligns with the evidence of Swinkels (2023), who utilizes a smaller sample. The maxima of both *HHI** can be attributed to a not fully transferred project with a single investor who sold off large parts of the investment after the STO. Apart from the maxima, the overall distributions remain the same, suggesting that secondary market trading does not change the ownership structure. Digitized properties are held in the mean by 401.2 different investors. Even though we obverse extreme cases, such as one property in 1,173 wallets, this variable is affected by the amount of *Total Investment*, since most issued tokens amount to around 50 USD and a higher *Total Investment* enables more investors to invest in a particular property.

The analysis of blockchain transactions on the buy side in Panel C shows that investors spend 317.82 USD in the mean during the STO and a median amount of 57.96 USD, which approximately equals the value of one token. With a mean *Secondary Market Buy* amounting to 88.70 USD, investors appear to acquire tokens mainly during the STO, while secondary market purchases play a subordinate role. This finding is underpinned by the ratio *Secondary Market Buy/Existing Exposure*, which indicates that investors raise their investment by a median value of 3% on the secondary market compared to their existing exposure.

Lastly, in Panel D, we examine the sell side. The *Secondary Market Sell* has a mean value of 99.97 USD. However, there exists a disparity in the number of observations between *Secondary Market Buy* and *Secondary Market Sell* due to the calculation of the daily balance, as explicated in Section 4.3.2. At the transaction-based level, each buy transaction corresponds to a sell transaction facilitated by the blockchain. Notably, in certain instances, a single sell transaction is associated with multiple buy transactions from different wallets (refer to transactions between wallet #2 and wallet #3 in Figure A.1 in the Appendix). Consequently, we observe a lower number of *Secondary Market Sell* observations compared to the number of *Secondary Market Buy* observations, alongside a higher mean value for *Secondary Market Sell*. Additionally, the distribution of *Secondary Market Buy* and *Secondary Market Sell* exhibits similarity from the maximum to the 25th percentile.[12]

[12]This observation suggests that the aggregation of sell transactions occurs within the range of transaction volumes below 3 USD (roughly the 25th percentile of both variables). To substantiate this claim, we conducted an unreported analysis utilizing kernel density plots for *Secondary Market Buy* and *Secondary Market Sell* observations across different transaction volume ranges. The plots reveal a large overlap between the two distributions. Consequently, the mean value of *Secondary Market Sell* is influenced by

The ratio *Secondary Market Sell/Existing Exposure* reveals that, in the mean, 9% of the existing exposure is sold, while the median value is 2%. The latter two variables highlight that most real estate token investors tend to hold their tokens and do not liquidate the investment quickly. The *Holding Period Sellers* shows that investors who sell their tokens hold them for 105.09 days in the mean before. This result is also consistent with Auer and Tercero-Lucas (2022), who find evidence of the increasingly popular *"hodling strategy"* among crypto investors who buy-and-hold tokens for a long time to avoid exposure to the short-term volatility in the crypto market.

4.4.2 Analysis of STO success determinants

To test our hypotheses for STO success, we run different regression specifications for the two success variables: funding time and speed. First, we sub-categorize funding time into *Funding Time until Success* for the successfully funded projects with OLS regressions (Models 1-2) and *Funding Time until Dec 2021* for all projects with parametric accelerated failure-time survival models with a lognormal distribution (Model 3), and loglogistic distribution (Model 4). We report the results in Table 4.3.

In the block of property characteristics for Hypothesis 1, only *Single Family* is positively related to the funding time of successfully funded and all projects. Based on the regression estimations, we find that *Single Family* increases the funding time of successfully funded projects by over 20 days in Models 1-2 and delays the success by around 79% ($e^{0.58} - 1$) for all projects in Model 4. The coefficients of *Detroit* and *Age* are significant for all projects in Model 4 and delay the success by 256% and 1%, respectively. Thus properties outside of *Detroit* – a city suffering from an enduring economic decline and shrinking population – are funded more quickly for reasons of diversification. In sum, we find supportive evidence in favor of Hypothesis 1 for funding time, i.e. that, the variables *Single Family*, *Detroit*, and *Age* are positively related to the success of an STO. However, since the remaining property-specific variables *Lot Size*, *Section 8*, and *Distance DTWN* are insignificant in all model specifications, we cannot provide further empirical support for Hypothesis 1. Particularly interesting is the irrelevance of the factors of size and location, which are typically important predictors in the real estate sector.

The campaign variable *#Pictures* is insignificant in all four models.[13] Therefore, we cannot provide empirical evidence for Hypothesis 2 and the common finding in CF that a more detailed description reduces information asymmetries and, hence, increases project success. The reason for this could be that, in contrast to conventional CF, in which information asymmetries are high (Courtney et al., 2017), the quality of a property can be determined

fewer observations.

[13] We do not anymore consider *#Characters* in the multivariate analysis, as outlined in Subsection 4.3.5; however, we find in unreported analysis that it is also insignificant.

Table 4.3: Determinants of Funding Time

	Dependent variable:			
	Funding Time until Success		Funding Time until Dec 2021	
	OLS		*AFT*	
			lognormal	*loglogistic*
	(1)	(2)	(3)	(4)
Rent per Token p.a.	10.05***	9.92***	0.20**	0.19**
	(2.71)	(3.15)	(2.00)	(2.20)
Expected Yield	−1,392.25	−519.07	−57.31***	−67.48***
	(−1.53)	(−0.55)	(−3.18)	(−3.72)
Total Investment	−0.004	−0.03	0.001	0.001*
	(−0.13)	(−0.91)	(1.47)	(1.78)
Age	0.05	−0.07	0.01	0.01*
	(0.17)	(−0.27)	(1.08)	(1.74)
Lot Size	0.002	0.002	−0.0000	0.0000
	(1.16)	(1.13)	(−0.32)	(0.35)
Section 8	−13.04	−3.56	0.12	−0.06
	(−1.45)	(−0.38)	(0.39)	(−0.22)
Single Family	24.84**	21.68**	0.47	0.58**
	(2.28)	(2.24)	(1.53)	(2.00)
Distance DTWN	0.83	0.70	0.01	0.002
	(0.39)	(0.34)	(0.12)	(0.05)
Detroit	7.10	5.03	1.05***	1.27***
	(0.46)	(0.36)	(3.21)	(3.83)
#Pictures	−0.89	−0.09	−0.03	−0.01
	(−0.56)	(−0.06)	(−0.96)	(−0.37)
Gas Fees		3.15**	0.10***	0.09***
		(2.27)	(4.04)	(3.56)
Housing Market		0.54**	0.01	0.01
		(1.96)	(0.88)	(0.78)
Constant	126.00	−44.36	6.53**	7.02**
	(1.30)	(−0.36)	(2.26)	(2.50)
Quarter-Year FE	Yes	Yes	Yes	Yes
Observations	122	122	164	164
R^2	0.48	0.52	/	/
Adjusted R^2	0.38	0.42	/	/
Log Likelihood	/	/	−577.14	−573.64
χ^2 (df = 21)	/	/	178.48***	193.30***

Note: The table reports the results for the sub-sample of successfully funded STOs with the dependent variable *Funding Time until Success* in Models 1-2 estimating OLS regression with robust standard errors. Models 3-4 present the results of the Accelerated Failure-Time (AFT) models with a lognormal and loglogistic distribution for all STOs, including unsuccessful ones with the dependent variable *Funding Time until Dec 2021*. The table contains the coefficient estimates and the corresponding *t*-statistics; the coefficients for the AFT model need to be exponentiated to interpret them as time ratios. All of the models include quarter-year dummies for time fixed-effects. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1 in the Appendix.

more easily. Thus information asymmetries are, in general, lower for real estate tokens than for CF projects.

The coefficient of *Gas Fees* is significant and positively related to both sub-categories of funding time. For example, higher transaction costs delay the success by around 9% for all projects in Model 4. This finding aligns with Momtaz et al. (2022), who find that investors limit their token trading activity when transaction costs are high. In sum, we find supportive evidence for Hypothesis 3b that investors reduce their trading activity when blockchain-related demand-driven transaction costs increase, which makes real estate STOs less successful.

The *Housing Market* coefficient is only significant and positively connected to *Funding Time until Success* in Model 2. However, funding time positively correlates with *Total Investment* and, as both *Total Investment* and *Housing Market* increase in our sample over time, we observe a positive coefficient for *Housing Market*. Among the financial controls, *Expected Yield* is significant for all projects and decreases the funding time strongly since a higher *Expected Yield* makes a project more attractive for investors. In contrast, *Rent per Token p.a.* positively impacts the funding time in all models. This result emanates from the fact that *Rent per Token p.a.* is in the same range for most observations due to the setting of the token issuer; however, just a few STOs above the 75% percentile (see Table 4.1) have not been successful and are the reason for the counterintuitive direction of effect of the *Rent per Token p.a.* coefficient. The *Total Investment*, which is significant for all projects with a loglogistic distribution, delays the success by merely 0.1%.

Models 1-2 consider only successful projects, and the estimations could be subject to a survivorship bias. However, comparing the results of the models of the successfully funded projects (Models 1-2) with those of all projects (Models 3-4), we do not observe apparent differences in signs and significances of the coefficients that would indicate a bias. The results of the two AFT models with different distribution assumptions are similar.

To obtain the complete picture of STO success and to rule out effects caused by the magnitude of the *Total Investment* amount, we study the newly-constructed dependent variable speed and present the results in Table 4.4. Since the STO is more successful if it raises more money within a certain period, the signs' interpretation of the coefficients should be opposite to the previous analyses of the funding time. Again, we sub-categorize the dependent variable into *Speed until Success* in Models 1-2 and *Speed until Dec 2021* in Models 3-4 and run OLS regressions.

Lot Size and *Detroit* are significant variables within property characteristics in all models for the speed variables. *Lot Size* is positively associated with both speed variables. Properties in *Detroit* have a lower *Speed until Success* of 26,080 USD/day for successful projects and a lower *Speed until Dec 2021* of 13,260 USD/day for all projects. In line with the traditional

Table 4.4: Determinants of Speed

	Dependent variable:			
	Speed until Success		Speed until Dec 2021	
	(1)	(2)	(3)	(4)
Rent per Token p.a.	−3.60	−3.94*	−2.84	−3.92*
	(−1.60)	(−1.74)	(−1.32)	(−1.75)
Expected Yield	656.78*	737.05	481.33	666.49*
	(1.71)	(1.43)	(1.53)	(1.73)
Age	−0.04	−0.03	0.004	0.07
	(−0.36)	(−0.25)	(0.04)	(0.61)
Lot size	0.003**	0.003**	0.002*	0.003**
	(2.00)	(2.12)	(1.83)	(2.19)
Section 8	−2.79	−3.58	0.19	−1.91
	(−0.54)	(−0.72)	(0.04)	(−0.44)
Single Family	−2.96	−2.34	−0.84	−1.41
	(−0.64)	(−0.53)	(−0.24)	(−0.42)
Distance DTWN	−1.28	−1.19	−0.24	−0.18
	(−1.51)	(−1.47)	(−0.31)	(−0.22)
Detroit	−28.70***	−26.08***	−16.79***	−13.26**
	(−3.30)	(−3.02)	(−2.77)	(−2.36)
#Pictures	0.63	0.57	0.88	0.45
	(0.78)	(0.71)	(1.34)	(0.73)
Gas Fees		−0.92**		−1.24***
		(−2.13)		(−3.14)
Housing Market		0.12		0.18
		(0.76)		(1.29)
Constant	−19.75	−43.90	−31.76	−66.76
	(−0.42)	(−0.54)	(−0.71)	(−1.08)
Quarter-Year FE	Yes	Yes	Yes	Yes
Observations	122	122	164	164
R^2	0.59	0.61	0.43	0.48
Adjusted R^2	0.51	0.53	0.36	0.41

Note: The table reports the results for the sub-sample of successfully funded STOs with the dependent variable *Speed until Success* in Models 1-2 and for the whole sample with *Speed until Dec 2021* in Models 3-4 estimating OLS regression with robust standard errors. The table reports the coefficient estimates and the corresponding t-statistics; all of the models include quarter-year dummies for time fixed-effects. The dependent variable *Speed until Success* is the fraction of *Total Investment/Funding Time until Success* and *Speed until Dec 2021* is the fraction of *Total Investment/Funding Time until Dec 2021.* The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1 in the Appendix.

real estate literature on location, this determinant is relevant, particularly for successfully funded projects. Since the majority of property characteristics are insignificant, we find only statistical support in favor of Hypothesis 1 for *Lot Size* and *Detroit*.

The campaign variable *#Pictures* is also insignificant for the speed variables.[14] The reason for this is probably the same as outlined above for the funding time. Consequently, we find no empirical evidence for Hypothesis 2.

We find a significant and negative relationship between the transaction costs *Gas Fees* and both speed variables, indicating that higher transaction costs are related to a lower level of STO success. For example, a one-standard-deviation increase in *Gas Fees* is associated with a 5,617 USD/day decrease in the *Speed until Dec 2021*. Compared to Model 2, the effect is more pronounced in terms of significance and magnitude of the coefficient for Model 4, which considers the whole sample. This finding is reasonable because this specification additionally considers unsuccessful projects whose success is more negatively affected by high transaction costs. Thus we find strong empirical support for Hypothesis 3b.

As assumed after taking the *Total Investment* into account for the dependent variable, *Housing Market* is insignificant. Among the financial characteristics, *Rent per Token p.a.* again has a negative impact in Models 2 and 4. The coefficient of *Expected Yield* is significant and positive on the 10% level in Models 1 and 4 and highly increases the speed of funding.

The adjusted R^2 ranges from 0.36 to 0.53. In summary, we observe that concerning both speed sub-categories, the traditional property characteristics of size and location (*Lot Size* and *Detroit*) are relevant determinants of STO success in addition to transaction costs on the crypto market (*Gas Fees*) and financial controls. The coefficient of *Lot Size* has the same magnitude for all models, *Detroit* shows a larger effect when restricting to only successfully funded projects. The unattractive location of the city of *Detroit* reduces the speed for successfully funded projects. *Gas Fees* is the only variable with a stronger effect when considering the entire sample, including unsuccessfully funded projects whose success is more negatively affected by high transaction costs. In line with Table 4.3, we do not observe apparent differences in the signs and significances between the models relying only on successfully funded projects and those comprising all projects.

4.4.3 Analysis of funding determinants

In the following, we study the funding determinants to analyze the entire crypto market on the macro-level and to account for its particularities over time. In Model 1 of Table 4.5, we present the regression estimations for the dependent variable daily *Inflows* per property

[14]The same applies if we include *#Characters*.

Table 4.5: Funding Determinants

	Dependent variable:	
	Inflow	Outflow
	(1)	(2)
ETH Price	139.72***	2.65
	(3.39)	(0.52)
Gas Fees	−1.28***	−0.10***
	(−11.04)	(−6.78)
ETH Shock	−607.06***	−33.96
	(−2.81)	(−1.29)
Gas Shock	−489.47**	49.25*
	(-2.20)	(1.81)
One-month Treasury	1,202.02*	64.15
	(1.86)	(0.73)
Ten-year Treasury	315.63	−28.49
	(0.51)	(−0.37)
ADS Index	−32.89	−8.99**
	(−0.90)	(−2.01)
Individual FE	Yes	Yes
Time FE	Yes	Yes
Observations	18,182	17,606
R^2	0.062	0.049
Adjusted R^2	0.053	0.040

Note: This table presents the analysis of funding determinants based on OLS regressions. It reports the coefficient estimates and the corresponding t-statistics. The dependent variable is either daily *Inflow* or daily *Outflow* per property in a fixed-effects panel regression with individual and time-fixed effects. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1 in the Appendix.

from investors purchasing tokens. Model 2 exemplifies the daily *Outflows* per property from investors selling tokens.

At first, we analyze the determinants that relate specifically to the crypto market. Model 1 exhibits a significant and negative coefficient of the *ETH Price* for *Inflows*, and no significance for *Outflows*. An increase of 1 USD in the *ETH Price* is associated with an increase of 139.72 USD in daily *Inflows* per property. Consequently, the crypto market sentiment appears to be a relevant predictor for capital *Inflows* on the market for real estate tokens, probably because crypto investors are subject to herding behavior caused by the sentiment on the crypto market. The results of *ETH Price* for *Inflows* provide statistical support for Hypothesis 3a, whereas we find no evidence of *Outflows* for Hypothesis 3a. Further, the coefficients of *Gas Fees* are negatively related to both capital *Inflows* and *Outflows*. The results of *Gas Fees* are consistent with Hypothesis 3b, that investors limit their trading activity to avoid high transaction costs, regardless of whether *Inflows* or *Outflows* are considered. It is worth noting that the crypto market sentiment *ETH Price* is not significantly related to *Outflows*, but crypto market transaction costs are. The reason for this could be that investors who have already decided to liquidate the tokens are timing the sale depending on transaction costs, as these directly affect their return on investment. Both dummy variables for short-term shocks on the crypto market are significant and negatively associated with *Inflows*, although with low or no significance for *Outflows*. To be more precise, the coefficient of *ETH Shock* decreases *Inflows* for 607.06 USD when the cumulative Ether return decreased for five days prior to the *Inflow*. The effect for a *Gas Shock* is less pronounced and implies that the occurrence of a *Gas Shock* decreases *Inflows* by 489.47 USD. The shock results for *Inflows* align with our crypto-market related Hypotheses 3a and 3b since a shock of the crypto market sentiment and the transaction costs reduce *Inflows*. Interestingly, short-term shocks in the crypto market do not seem to play a major role in *Outflows*. Possibly this is because regular cash flows from the tokens are based on rent payments and are not affected by short-term crypto shocks, so there is no incentive to sell and cause an *Outflow*. Consequently, we cannot provide empirical evidence for *Outflows* and the shock variables for our Hypotheses 3a and 3b.

Regarding the macroeconomic factors *One-month Treasury*, *Ten-year Treasury*, and the *ADS Index*, we find occasional and low significances for both *Inflows* and *Outflows*. The short-term interest rate has a positive and significant influence on *Inflows*, whereas long-term interest is insignificant for both capital flows. An increase in the *ADS Index*, indicating progressively better-than-average conditions for doing business, significantly reduces *Outflows*. Thus the macroeconomic situation does not appear to be an essential criterion in the decision-making process of a real estate token investor. Our finding is consistent with Yermack (2015) and Bianchi (2020), who conclude that macroeconomic events and factors do not drive trading volumes and daily exchange rates of the main cryptocurrencies.

In sum, we find that the crypto market-related transaction costs, sentiment, and the corresponding short-term shocks are relevant predictors of daily *Inflows* for purchasing tokens rather than daily *Outflows* of selling tokens.

4.5 Robustness and further analysis

4.5.1 Adjustment of financing threshold

It is common practice that RealT retains around 5% tokens of a property to ensure liquidity on secondary markets in the future, which is why we define the success of a project as transferring 95% of the tokens. We vary the threshold for the definition of "successfully" funded between 90% and 100% in unreported analyses. Our results remain qualitatively unchanged and robust for these adjustments.

4.5.2 Analysis of the determinants of Total Investment and Expected Yield

Digging deeper into the structure of the projects offered in the STO, we investigate the determinants of the money-oriented variable *Total Investment* and present the estimations in Models 1-2 in Table A.3 in the Appendix. Regarding the financial variables, the coefficient of *Expected Yield* is significant and negative in both model specifications. When considering the property characteristics, we find that lower quality properties, which are older and have higher risk diversification among tenants, are offered with a lower *Total Investment.* The variables *Lot Size* and *Section 8* have a significant positive impact across all models. The lower risk of a rent default of *Section 8* supported rents is associated with a higher *Total Investment.* The coefficient of the CF variable *#Pictures* is insignificant, probably because this variable is less relevant to the token issuer. Both market-related variables *Gas Fees* and *Housing Market* are insignificant. In the next step, we switch from the dollar amount of *Total Investment* to a return perspective and study the determinants of *Expected Yield* in Models 3-4 in Table A.3. As expected, the *Rent per Token p.a.* is positively related to the *Expected Yield.* In line with the previous results for *Total Investment*, the coefficient for *Single Family* is also negatively related to *Expected Yield.* The coefficient of *Distance DTWN* indicates that a higher distance from downtown reduces the yield due to lower rent in more unattractive locations further afield. The *Housing Market* is negatively associated with the *Expected Yield.* A higher housing index is connected with higher housing and token values and, consequently, a lower *Expected Yield.*

In summation, only for *Single Family* do we find consistent signs and significances for both *Total Investment* and *Expected Yield*, while the evidence for the remaining variables is mixed. While crypto-market transaction costs are significantly related to the success of the STO as measured by funding time and speed, see Subsection 4.3.3, they are not related to the *ex-ante* set structure of the offered projects by the token issuer.

4.6 Conclusion

Digitization is transforming various industries, including the financial and real estate sectors. We highlight the new way of securitizing assets, using the blockchain and digital security tokens and their issuance processes through STOs. Real estate has been identified as a suitable market for tokenization due to this technical innovation overcoming the drawbacks of direct real estate investments, such as high entry barriers and illiquidity. Technical features facilitate the investment of small amounts of money, eliminate the need for financial intermediaries, and increase transaction speed, consequently lowering the costs for all parties involved. Thus investors can diversify their portfolios more easily among asset classes and countries. The tokens can be traded after issuance on secondary markets, which enables liquidity. Even though the possibility of fractional ownership already exists in indirect investment instruments, such as funds or REITs, real estate tokens come closer to direct ownership with controlling rights.

Based on STO data of 173 real estate tokens and more than 238,433 blockchain transactions, we analyze investor behavior, the determinants of STO success, and capital flows over time. During our observation period, real estate token investors hold a mean of 10 different tokens and an investment amount of 4,030 USD, which shows that investors do not yet hold well-diversified real estate token portfolios. Ownership of the properties is not concentrated on some large investors emphasizing that tokenization provides broad access to real estate ownership for many small investors. Further, we conclude that investors acquire tokens mainly during the STO, while the secondary market plays a subordinate role in token purchases and sales. This study's primary success variable of interest is the mean funded investment amount per day (*Speed*). Property-specific fundamentals and crypto market-related transaction costs are positively related to STO success, along with financial characteristics. In line with the well-known explanatory power of location factors in real estate, we find that location is another important determinant of STO success. The success of STOs appears to be independent of crowdfunding characteristics, probably because a property's quality can be determined more easily, and information asymmetries are lower than for conventional crowdfunding projects. Investors seek diversification possibilities through location choice to reduce the idiosyncratic cash flow risk of the investment and try to evade high transaction costs that reduce their return. From the perspective of capital inflows (token purchases) and capital outflows (token sales) per day, we find that real estate token investors pay equal attention to the crypto market-specific sentiment and transaction costs when purchasing tokens. In contrast, only the transaction costs directly reducing the return on investment are relevant for sales. Both short-term shocks have a strong negative impact on capital inflows. Macroeconomic factors appear to have little effect on capital flows in general. These results highlight the importance of considering the specific crypto market environment and the characteristics of the underlying asset class for real asset tokenization.

A limitation is our small sample size of 173 projects, resulting from the fact that tokens are becoming the focus of public attention. Our results may not be generalized, as they are derived from observing a small but growing number of crypto enthusiasts familiar with the technical background. Therefore, there is an avenue for future research to test and verify our results in a broader sample regarding other asset classes, periods examined, geographic scope related to different jurisdictions and implementation options, and the number of investors.

Our study has practical and policy implications. As discussed at the G-7 meeting in May

2022, various regulators and politicians have called for accelerating global crypto regulations for better financial stability to enable innovative digital finance solutions and investor protection. Our findings contribute to the last two objectives of this regulatory effort. We find that the particularities of the crypto market are essential determinants for the success of real estate STOs and capital flows. This result may raise the concern that token investors mainly follow trends that do not reflect the fundamental asset characteristics, implying a high need for consumer protection. Such technical innovation can also support investors in building more diversified portfolios. However, according to our results, this possibility has not been used sufficiently until now. Regulators must find a compromise to achieve investor protection and foster the development of digital finance products without suppressing the opportunities for technology and innovation.

Appendix

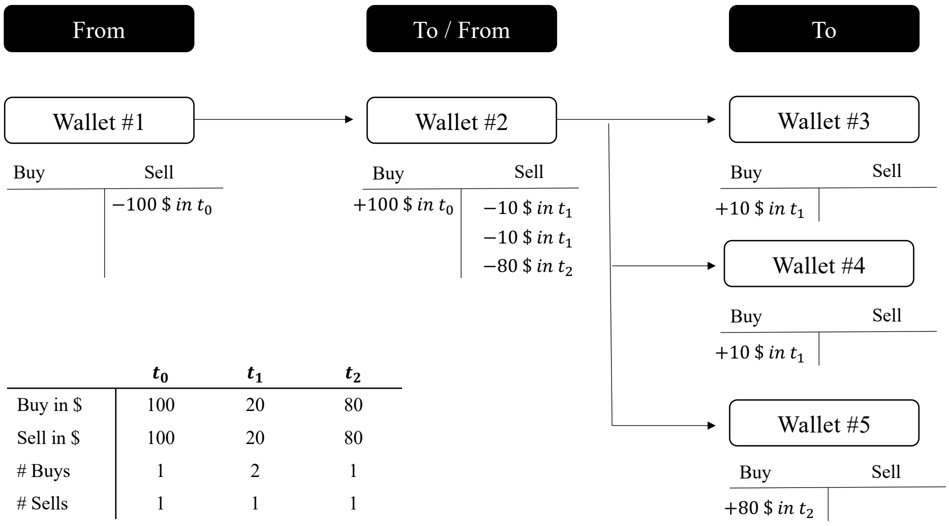
Table A.1: Definition of all Variables

Blockchain transaction analysis		
Properties per Investor	Number of distinct real estate tokens per unique wallet	Own calculations
Holdings per Investor as of Dec 2021	Dollar Holdings per Investor as of 31 Dec 2021	Own calculations
Holding Period all Investors as of Dec 2021	Holding Period of all Investors in Days as of 31 Dec 2021	Own calculations
HHI	Herfindahl-Hirschman Index per property	Own calculations
HHI STO*	Normalized Herfindahl-Hirschman Index per property after the tokens have been transferred during the STO, based on the actual quantity of issued tokens comprising successful and unsuccessful STOs (between 0 and 1)	Own calculations
HHI as of Dec 2021*	Normalized Herfindahl-Hirschman Index per property as of 31 Dec 2021 (between 0 and 1)	Own calculations
Investors per Property	Number of unique wallets per real estate token	Own calculations
STO Buy	Amount of money of buy transactions during the STO	Own calculations
Secondary Market Buy	Amount of money of secondary market buy transactions in USD	Own calculations
Secondary Market Buy/ Existing Exposure	Percentage ratio of the *Secondary Market Buy* to the existing exposure	Own calculations
Secondary Market Sell	Amount of money of secondary market sell transactions in USD	Own calculations
Secondary Market Sell/ Existing Exposure	Percentage ratio of the *Secondary Market Buy* to the existing exposure	Own calculations
Holding Period Sellers	Holding Period of investors selling tokens in days	Own calculations
Analysis of STO determinants		
Dependent variables		
Funding Time until Success	Number of days until all tokens (95 percent, since RealT keeps around 5 percent to themselves) are transferred to wallets. For this variable, only successful projects are considered. The start date of the funding period is derived from the HTML code on the website and the end date from the blockchain explorers.	Own calculations
Funding Time until Dec 2021	Number of days until all tokens (95 percent, since RealT keeps around 5 percent to themselves) are transferred to wallets. For this variable, both successful and unsuccessful projects are considered. The start date of the funding period is derived from the HTML code on the website and the end date from the blockchain explorers.	Own calculations
Speed until Success	95% of *Total Investment* divided through *Funding Time until Success*, (in thousands USD/day) for the sub-sample of successful projects	Own calculations
Speed until Dec 2021	95% of *Total Investment* divided through *Funding Time until Dec 2021* (in thousands USD/day) for all projects. For projects that have not been successfully funded until the end of our observation period and are on the market longer than the mean time of *Funding Time until Success*, the *Speed until Dec 2021* is equated to 0. For projects that have not been successfully funded until the end of the observation period and are on the market shorter than the mean time of	Own calculations

	Funding Time until Success, the actual amount of money raised is used instead of *Total Investment.*	
Explanatory variables		
Rent per Token p.a.	Rent per token per year	RealT
Total Investment	Amount of money required for the funding, technically the number of tokens multiplied by the token price (in thousands USD)	RealT
Expected yield	Expected income calculated as net rent divided by token price	RealT
Age	Difference between the publication date of the project and the construction year	RealT
Lot Size	Size of the real estate (in square foot)	RealT
Section 8	Percentage of rents supported by Section 8 in the whole property	RealT
Single Family	A dummy variable for the property type of use that shows whether the building is a single-tenant property, 0 otherwise.	RealT
Distance DTWN	Distance to downtown in miles	Walk Score
Detroit	A dummy variable that shows whether the property is located in Detroit, 0 otherwise.	RealT
#Pictures	Absolute numbers of pictures of the property published on the platform	RealT
#Characters	Absolute number of characters of the descriptive text of the project on the platform	RealT
Gas Fees	Transaction costs on the Ethereum blockchain on the day the project is published online or on the day of the observation, converted to USD	Coinmarketcap
Housing Market	S&P Case-Shiller Home Price Index for the corresponding region, lagged for one month	S&P Dow Jones Indices
Analysis of funding determinants		
Dependent variables		
Inflow	Daily capital inflows per property per day in USD (STO and secondary market buy transactions)	Own calculations
Outflow	Daily capital outflows per property per day in USD (secondary market sell transactions)	Own calculations
Explanatory variables		
One-month Treasury	Market yield on US Treasury Securities at 1-month constant maturity, quoted on an investment basis	FRED, Federal Reserve Bank of St. Louis
Ten-year Treasury	Market yield on US Treasury Securities at 10-year constant maturity, quoted on an investment basis	FRED, Federal Reserve Bank of St. Louis
ADS Index	Aruoba-Diebold-Scotti (ADS) Business Condition Index based on Aruoba et al. (2009) to measure macro-economic activity at a daily frequency	Federal Reserve Bank of Philadelphia
ETH Price	Cumulative return of Ether over a period of five days before the observation	Coinmarketcap
ETH Shock	A dummy variable that equals one if the cumulative return of *ETHPrice* decreased by more than 5% over a five-day window before the observation, 0 otherwise.	Own calculations
Gas Shock	A dummy variable that equals one if the *Gas Fees* cumulatively increased by more than 5% over a five-day window before the observation, 0 otherwise.	Own calculations

Note: List and definitions of all variables and the corresponding sources. RealT as a source corresponds to information obtained from RealToken's website.

Figure A.1: Blockchain Analysis Scheme



	t_0	t_1	t_2
Buy in \$	100	20	80
Sell in \$	100	20	80
# Buys	1	2	1
# Sells	1	1	1

Note: This Figure illustrates the calculation scheme for determining the buy and sell sides in Table 4.2. We used a daily balance calculation for each wallet to evaluate changes over time. An increase in balance represents a buy transaction, while a decrease indicates a sell transaction.

Table A.2: Correlation Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) Total Investment	1													
(2) Expected Yield	−0.03	1												
(3) Funding Time until Success	0.01	0.06	1	1										
(4) Funding Time until Dec 2021	0.01	0.06	1	1										
(5) Speed until Success	0.43	0.01	−0.37	−0.37	1	1								
(6) Speed until Dec 2021	0.43	0.01	−0.37	−0.37	1	1								
(7) Rent per Token p.a	0.26	0.42	0.29	0.29	−0.04	−0.04	1							
(8) Age	0.12	0.03	−0.17	−0.17	0.24	0.24	−0.24	1						
(9) Lot Size	0.55	0.24	0.13	0.13	0.36	0.36	0.60	−0.17	1					
(10) Section 8	0.19	−0.19	−0.23	−0.23	0.08	0.08	0.06	−0.02	0.05	1				
(11) Distance DTWN	−0.08	−0.02	0.03	0.03	−0.13	−0.13	0.13	−0.02	0.08	0.03	1			
(12) #Pictures	0.30	0.05	−0.18	−0.18	0.38	0.38	0.14	0.11	0.32	−0.02	−0.01	1		
(13) Gas Fees	0.19	−0.03	0.39	0.39	−0.12	−0.12	−0.16	0.12	−0.01	−0.28	0.04	−0.11	1	
(14) Housing Market	0.34	−0.49	0.06	0.06	0.24	0.24	−0.28	0.29	−0.08	−0.09	−0.12	0.01	0.27	1

Note: This table reports the Bravais-Pearson correlation coefficients of the dependent and explanatory variables. All variables are defined in Table A.1 in the Appendix.

Table A.3: Determinants of Total Investment and Expected Yield

	Dependent variable:			
	Total Investment		Expected Yield	
	(1)	(2)	(3)	(4)
Rent per token p.a.	−1.61	−8.84	0.002**	0.002***
	(−0.11)	(−0.56)	(2.47)	(3.91)
Expected Yield	−6,629.80**	−4,471.23*		
	(−2.55)	(−1.77)		
Age	−2.74**	−2.31*	0.0001**	0.0000
	(−2.40)	(−1.82)	(2.33)	(0.60)
Lot Size	0.02***	0.02***	0.0000	−0.0000
	(2.84)	(3.02)	(1.57)	(−0.07)
Section 8	139.29**	149.49***	−0.002	−0.003
	(2.56)	(2.75)	(−1.27)	(−1.64)
Single Family	−240.42***	−238.69***	−0.004***	−0.003**
	(−5.74)	(−5.75)	(−2.73)	(−2.25)
Distance DTWN	−8.08	−6.60	−0.001**	−0.001***
	(−1.25)	(−1.04)	(−2.18)	(-2.82)
Detroit	−10.88	6.95	0.01***	0.004**
	(−0.22)	(0.13)	(3.30)	(1.99)
#Pictures	5.99*	3.67	−0.0002	0.0002
	(1.67)	(0.89)	(−1.01)	(0.65)
Gas Fees		3.11		−0.0001
		(1.19)		(−1.09)
Housing Market		1.42		−0.0002***
			(1.53)	(−4.17)
Constant	1,289.40***	887.10**	0.09***	0.11***
	(3.95)	(2.32)	(5.54)	(9.90)
Quarter-Year FE	Yes	Yes	Yes	Yes
Observations	165	165	165	165
R^2	0.58	0.70	0.65	0.66
Adjusted R^2	0.54	0.66	0.61	0.61

Note: This table presents the results of OLS regression for the dependent variables *Total Investments* and *Expected Yield* with robust standard errors. The table reports the coefficient estimates and the corresponding *t*-statistics; all models include quarter-year dummies for annually and quarterly fixed-effects. The dependent variable *Total Investment* is measured in thousands USD. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Table A.1 in the Appendix.

Conflicts of Interest

One author bought a few digital tokens issued by the company RealT so that they could describe the process of tokenization. The current value is lower than 200 USD. The other authors have no commercial relationship with the company or management whose data we mostly rely on.

Acknowledgements

We appreciate the helpful comments and suggestions from Laurens Swinkels, Andreia Dionisio, and other participants at the 2022 CryptoAssets and Digital Asset Investment Conference of the Future Finance and Economics Association (Rennes), 2022 ERES Meeting (Milan), Cryptocurrency Research Conference 2022 (Durham), NTNU Conference 2022 (Trondheim), EFA Workshop on FinTech – Blockchain & Cryptocurrency (Barcelona), 5th Annual REALPAC/Toronto Metropolitan University Research Symposium, and 2023 ARES Annual Meeting and Conference (San Antonio, Texas).

Chapter 5

German FinTech companies: A market overview and volume estimates

This research project is joint work with Gregor Dorfleitner (University of Regensburg) and Ralf Laschinger (University of Regensburg). The paper has been published as:

Dorfleitner, G., Kreppmeier, J., Laschinger, R. (2023). German FinTech companies: A market overview and volume estimates. *Credit and Capital Markets*, forthcoming.

Abstract The FinTech market in Germany is a dynamic and growing field that is difficult to observe in its entirety. This report provides a hand-collected market overview of the FinTech market in Germany, as well as an application case in terms of volume estimates for the financing and asset management segments through December 2021. The data includes various verified characteristics of 978 unique companies that can be classified under the financial technology sector and operate in Germany. Each observation represents a company with 24 variables, including name, address, legal form, founders with corresponding LinkedIn accounts, registration number or company ID, assignment to FinTech segments and sub-segments, banking cooperation, URL address, local court, former name, operating status. We provide the description of the variables as well as a taxonomy to categorize FinTechs. The dataset contains both established companies and startups and presents valuable information for researchers, practitioners and also regulators.

Keywords FinTech, Germany, Start-Up, Financial Technology, Digital Finance, Entrepreneurship, Supervision

JEL G10, G20, G28, K20, L81, M13

5.1 Introduction

The importance and market volumes of FinTech companies (FinTechs) have been growing for a number of years, making FinTechs a very relevant subject in the academic context as well as for practitioners and regulators. Due to the predominantly digital nature of FinTechs, these companies are often only observable through their web presence. Likewise, they are not monitored by any regulator, at least not in the early stage, which is the reason why there have been few centralized captures or aggregated industry reports. This report is divided into two parts. First, we describe the companies and variables included in our aggregated German FinTech database as of December 2021. Using the German FinTech list by Dorfleitner et al. (2017) as a starting point, we have collected aggregate information on 978 FinTech-related companies that are or were active in Germany. Second, as an application case of the provided data, we present market volume estimates for the FinTech segments of financing and asset management until December 2021.

5.2 Data description

The data set is accessible on the Mendeley Data repository (Dorfleitner et al., 2022). The data can be downloaded from the URL: https://doi.org/10.17632/438ytjyzxk.2 in an open access format.

5.2.1 Data collection

Our data were acquired in the following manner. The starting point was the FinTech list of Dorfleitner et al. (2017). This list already consists of hand-collected data over the years 2015 and 2016. In a similar vein, we continuously collected data until December 2021 using specific and topic-related databases (Crunchbase, BvD Dafne, German Company Register, Trade Register Excerpts), FinTech and bank websites as well as with structural Google searches. The entries and properties, specifically the operating status, were checked in regular time-intervals throughout the collection process over the years. The aim of the collection procedure was to find and identify all relevant FinTechs operating in Germany with a structured approach. Different databases and websites were used to obtain an overview of the market. The dataset was repeatedly updated and verified throughout the years within this process. An association to the segment of operations was conducted. Through structured Google searches the operating status was checked.

5.2.2 Variables description

Table 5.1 shows the overview of all variables in the dataset and describes the type and content of each variable. Note that for some of the 978 FinTech companies, some variables have missing values, which are marked NA.

The classification of FinTechs into segments and subsegments is generally based on the taxonomy of Dorfleitner et al., 2017, pp. 6-10, which is displayed in Figure 5.1. In order to take account of more recent developments in the market, we are also including the subsegment “BigTechs” for the payment services of BigTechs companies such as Amazon Pay, ApplePay and Google Pay under the segment “Payments”. In addition, we assign FinTechs operating in the field of blockchain and distributed ledger technology to the "Blockchain and cryptocurrencies" subsegment, which is subordinate to the "Payments" segment, although not all of them have business activities related to payment services. Companies offering services in the field of "RegTech" (Regulatory Technology) are only considered if there is a clear intersection with financial services and thus FinTech. They are assigned to a (sub-) segment according to the specific service provided, this is in the case of our dataset mostly “Technology, IT and Infrastructure” with services e.g., to detect financial fraud or ID-based for KYC purposes.

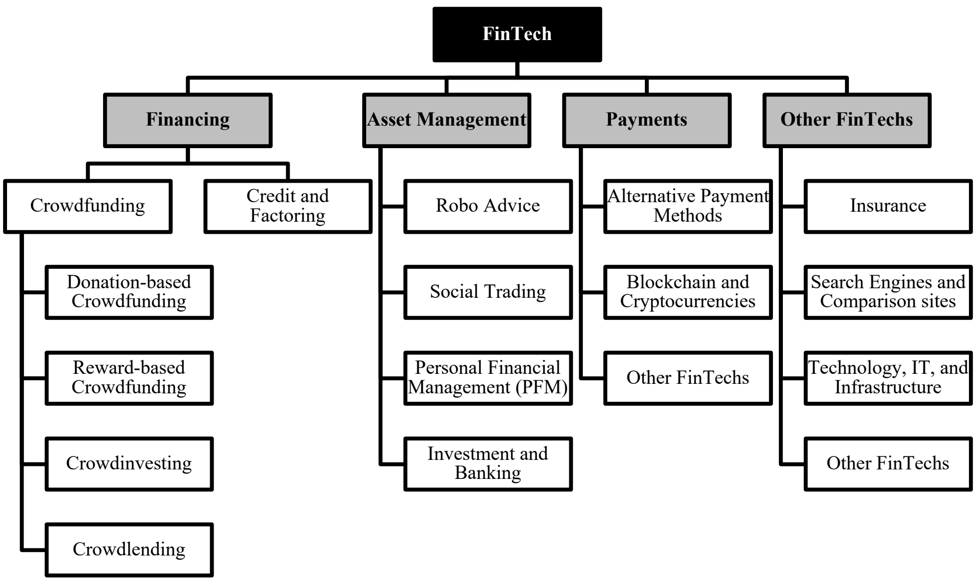


Figure 5.1: Taxonomy of FinTech companies according to Dorfleitner et al., 2017, p. 7

This dataset was created to identify all relevant FinTechs operating in Germany. Therefore, a structured approach was used combining different databases and websites, as listed above, to obtain and verify a possibly complete overview of the market. The dataset was repeatedly updated and verified throughout the years within this process. Furthermore, each FinTech was assigned to one (sub-)segment in which its main operations take place.

Table 5.1: Variables description

Variable	Type	Description
ID	Numeric	Unique identifier for each FinTech
Name	Character	Name of each FinTech
Status	Binary	FinTech is active up until 31.12.2021
Original German	Binary	1: FinTech is founded originally in Germany; 0: just operating in Germany
Founding year	Numeric	Year the FinTech was founded
Founder	Character	Name of the founder or founding company, either name of a natural person or company name, if several founders separated by ;
Founder (LinkedIn)	HTML	Link to the LinkedIn Profile, separated by ;
Legal Name	Character	Name of the FinTech according to company register/law
Legal Form	Character	Legal form of the FinTech according to law from company register
Street	Character	Street name of the FinTech according to the company register
Postal Code	Numeric	Postal code of the FinTech according to the company register
City	Character	City of the FinTech according to the company register
Country	Character	City of the FinTech according to the company register
Register Number / Company ID / LEI	Character	Register number / company ID / LEI of the FinTech
Segment /	Categorical	Association to an operating segment according to Fig. 1 below and description below (according to Dorfleitner et al., 2017)
Subsegment /	Categorical	Association to an operating subsegment according to Fig. 1 below and description below (according to Dorfleitner et al., 2017)
Bank Cooperation	Binary	1: There exists a cooperation with a private/commercial bank; 0 otherwise
Homepage	HTML	Homepage of the FinTech
E-Mail	Character	E-Mail address of the FinTech
Insolvency	Binary	1: FinTech is undergoing insolvency proceedings; 0 otherwise
Liquidation	Binary	1: FinTech has been liquidated; 0 otherwise
Date of Inactivity	Date	Date of cessation or date of opening insolvency proceedings or date of liquidation
Local court	Character	Local court in Germany of the FinTech, if the company is resident in Germany
Former name	Character	Former name(s) of the FinTech, if the company was renamed

Through structured Google searches the operating status was checked on a regular basis.

5.2.3 Descriptive statistics

Figures 5.2 and 5.3 show the number of companies identified in the various segments according to the taxonomy of Dorfleitner et al. (2017). It should be noted that there is no uniform distribution across the various segments. For example, at the end of 2021, most FinTechs are to be found in the payments segment with a number of 191, followed by the broad technology, IT and infrastructure segment with 127 companies. A progressive maturation of companies can be observed across all segments. At the same time, it should be emphasized at this point that the number of companies does not reflect the business volumes of the individual segments.

Figure 3 differentiates within the various segments based on the activity status of the FinTechs. The dataset also includes these inactive companies to ensure a survivorship bias-free dataset for further studies. The dataset contains an unknown number of companies that can still be reached via a website, but probably no longer have any business activity. Overall, it is noticeable that especially in the subsegments crowdinvesting and donation- and reward-based crowdfunding the highest shares of inactive companies were found. We also note that, in contrast to the venture capital industry, a large proportion of FinTechs are still active. Therefore, we additionally display in Figure 5.4 the average age per subsegment and differentiate between active and inactive FinTechs, whereby we can only calculate the age for 110 out of 172 inactive companies because of data availability. We cannot observe in any subsegment that the average age of active companies is close to that of inactive companies, which would explain the low number of inactive companies compared to the venture capital sector.

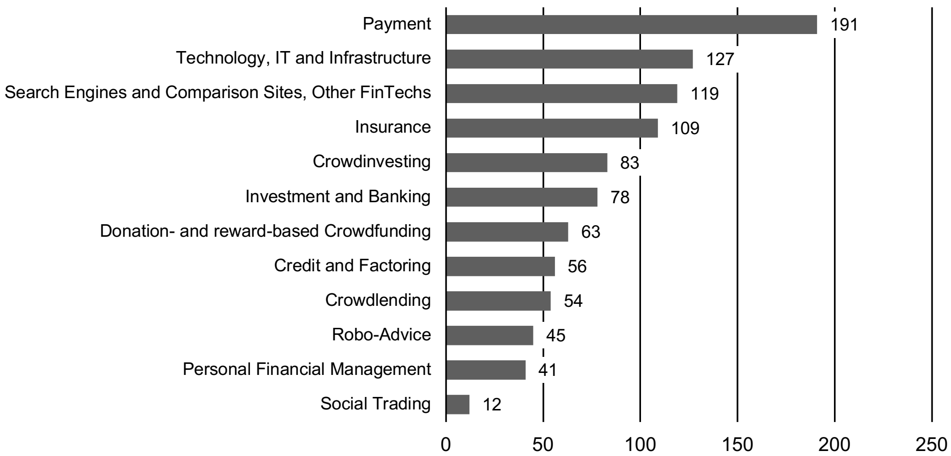


Figure 5.2: Absolute frequency of subsegments in our dataset

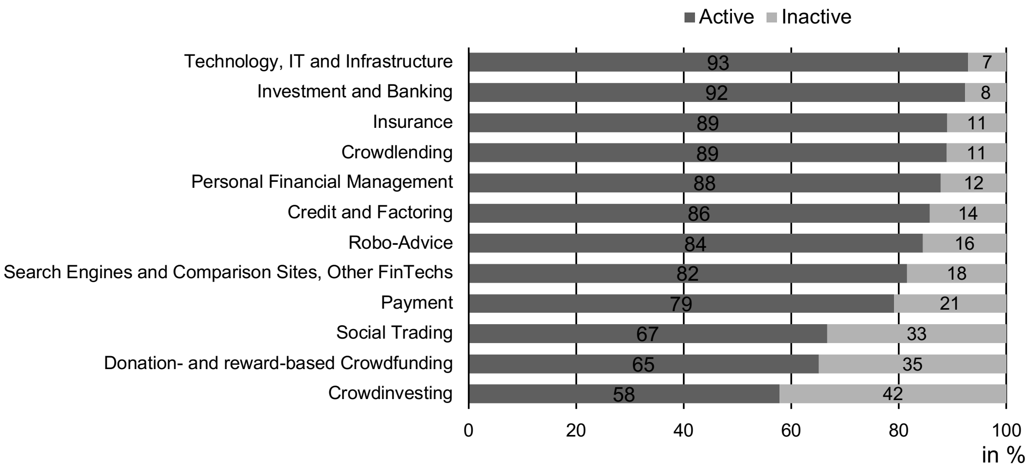


Figure 5.3: Relative frequency of active and inactive FinTechs in each subsegment

5.2.4 Previous use of the data in research

The first version of the dataset and the categorization is based on Dorfleitner et al. (2017). Afterwards, estimations for the German market volume were performed for several years and segments, see for instance Dorfleitner et al. (2020) and Dorfleitner and Hornuf (2023). Based on the observed German FinTech companies, empirical studies related to data protection and the General Data Protection Regulation matched with the privacy policies were performed with simple descriptives by Dorfleitner and Hornuf (2019) or with the help of textual data mining and in multivariate analysis by Dorfleitner et al. (2023).

5.3 Application case of the dataset: Estimation of market volumes of German FinTech segments

In this section, we present the estimation of the market volumes of German FinTechs as an application case for the dataset presented above. Based on the taxonomy of Dorfleitner et al. (2017), we focus on the financing and asset management segment. We exclude the payment segment as we do not have access to the transaction volumes of large players such as Paypal or ApplePay, which account for the majority of the market share in this segment. In addition, we exclude the Other FinTechs segment as for these companies data on market volumes cannot be collected in a comparable way.

To this end, we estimate the market volumes of 434 FinTechs, of which 341 are still active. To estimate the market volumes for the year 2021 in the each subsegment, we consider those three to five companies that had the highest market shares in 2020 and estimate their market volume in 2021 with the estimation and research techniques displayed in Dorfleitner

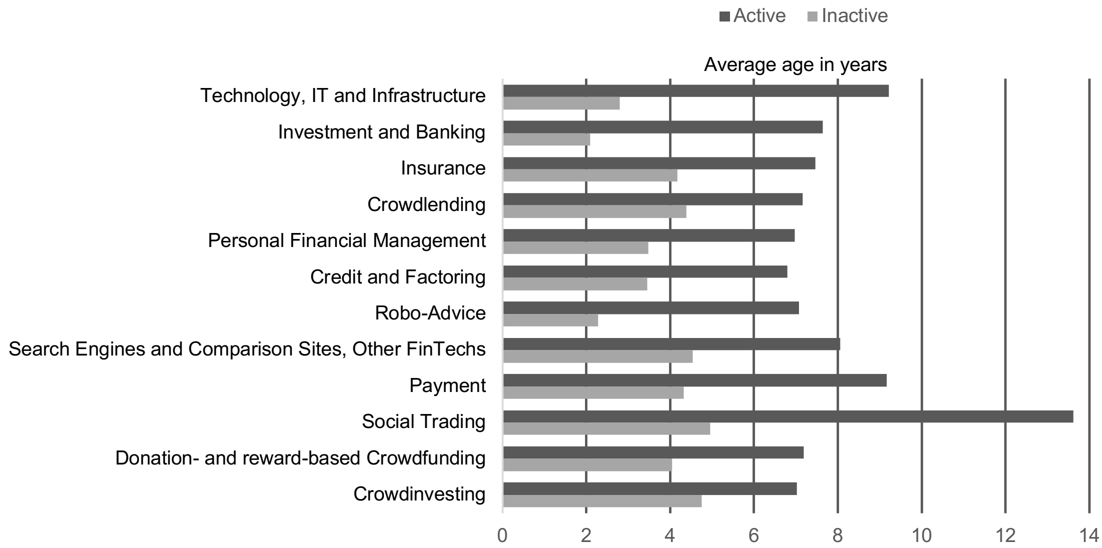


Figure 5.4: Age per subsegment in comparison for active and inactive FinTechs

et al. (2017), chapter 3, or Dorfleitner et al. (2020). The resulting relative market volume increase of those market leaders is then applied to the total 2020 subsegment figure as published by Dorfleitner and Hornuf (2023) in order to obtain a total market volume estimate for 2021.

Market volumes in all financing subsegments are supposed to represent *transaction volumes*, i.e. money raised, while market volumes in the Asset Management segment are meant to be value of money invested (in the sense of assets under management) by the FinTechs. Both specifications are in line with the mentioned literature, which addresses the same issue for the years before 2021.

Figure 5.5 presents the market volume development over time in the donation- and reward-based crowdfunding subsegment. In reward-based crowdfunding, investors receive a non-monetary consideration from the FinTechs for their financial support of a project which in many cases serves as a pre-financing of the products (Mollick, 2014). This can be of a purely non-material nature, for example in the form of a naming, but can also include material counter-values, such as the delivery of a product to be developed. Even if some platforms define a thematic focus, such as the mediation of regional, sustainable or sports-related projects, the intended use for the collected capital is often very different. Other platforms do not specialize in specific topics. Donation-based crowdfunding is characterized by the fact that the capital providers receive no or, in turn, only an ideal consideration for their financial contribution. Due to the operational overlap between the two subtypes of donation-based and consideration-based crowdfunding, the presentation of market volumes is summarized. While there still is a relative growth of roughly 20 per cent from 2020 to 2021, the absolute figures are still small. Nevertheless, this segment has seen significant growth during the covid-19 pandemic, as many individuals in their local area have supported small businesses, restaurants, bars, and cultural venues with donations. The German market leader which is originally German is the donation-based platform *Betterplace*, followed by *Startnext*, the largest non-original German platform *Kickstarter*, *Viele Schaffen Mehr* and *Indiegogo*.

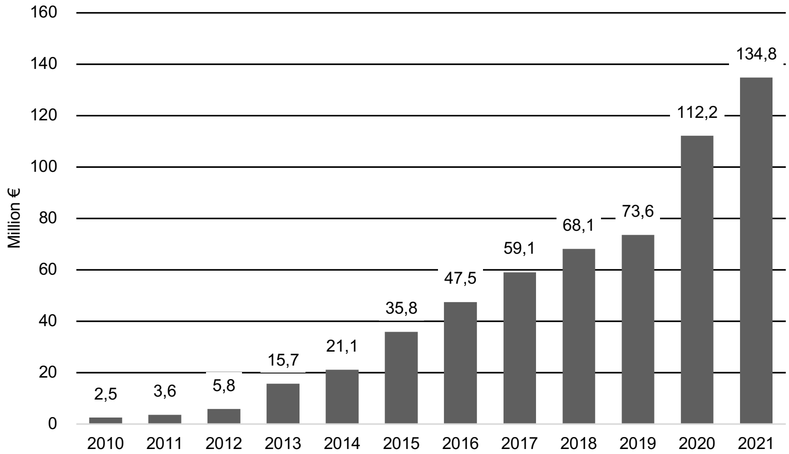


Figure 5.5: Market volumes of the subsegments donation- and reward-based crowdfunding over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2021)

Within the crowdinvesting subsegment, investors often receive an equity-like investment in the form of profit participation rights, dormant equity holdings, participatory loans or subordinated loans. They therefore participate financially in the future development of a company at the end of the term (Hainz et al., 2017, 2019). Note that, unlike in many other countries, in Germany crowdinvesting is not equity-based crowdfunding but rather financing through mezzanine forms such as junior debt. The market volume in the subsegment crowdinvesting (Figure 5.6) has experienced a decline in 2020 because of the covid-19 pandemic, which led to some distortions in the market. However, the crowdinvesting subsegment has recovered and reached an all-time high of 522,3 million EUR in 2021 with a growth rate of 40 per cent with respect to 2020. For crowdinvesting, the German market leader is *Exporo*, followed by *Bergfuerst*, *Companisto*, *Wiwin*, *SeedMatch*, *Zinsbaustein*, *Engel&Völkers*, *EstateGuru* and the non-German platform *Seedrs*.

The segment of crowdlending (Figure 5.7) is characterized by the fact that the capital providers receive predefined annuity payments immediately after financing in exchange for providing the financial resources. Investors and borrowers are either private individuals or companies. FinTechs merely act as intermediaries (Lee and Shin, 2018). The actual lending is handled by a partner bank. After a stagnation phase during the years 2018 until 2020 this segment sees now considerable growth. The market leader is the non-German platform *Loanboox* with approx. 2 billion to which the largest part of the growth in 2021

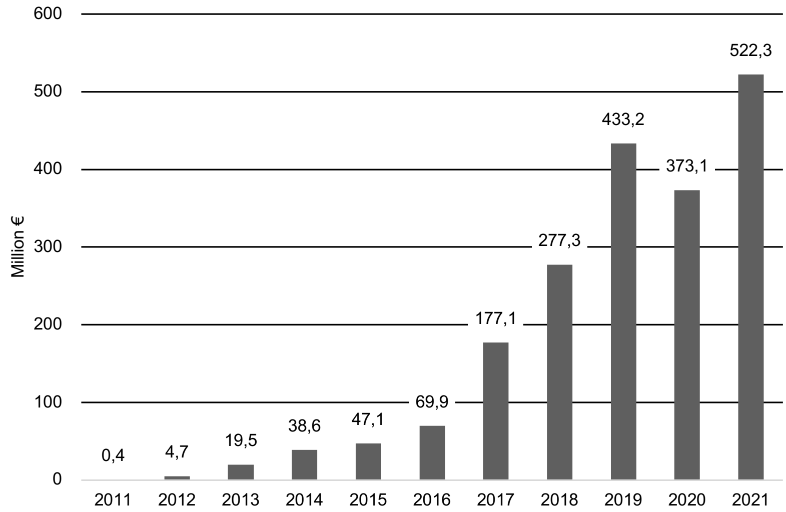


Figure 5.6: Market volume of the subsegment crowdinvesting over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2020 and 2021)

can be attributed, followed by the German platform *Auxmoney*, *Creditshelf* and the Latvian platform *Mintos*.

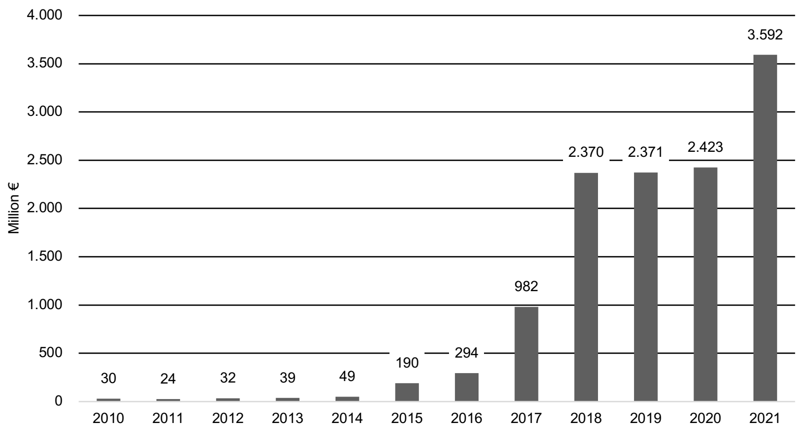


Figure 5.7: Market volume of the subsegment crowdlending over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2021)

Figure 5.8 now shows the aggregate volumes of the crowdlending, crowdinvesting and donation and reward-based crowdfunding segments with 4.249 billion, with crowdlending accounting for both the largest percentage share and the most dynamic growth.

The Credit and Factoring subsegment in Figure 5.9 includes FinTechs that act purely as an online alternative to traditional financing by a bank. Unlike the previous segments, however, the funds are not provided by the crowd. This form of financing is made available to both private individuals and companies (Dorfleitner et al., 2017). Different types of financing can be distinguished, such as traditional loans, online loans, installment loans, express loans or loans for financing the purchase of goods and credit-like factoring. Factoring, in particular,

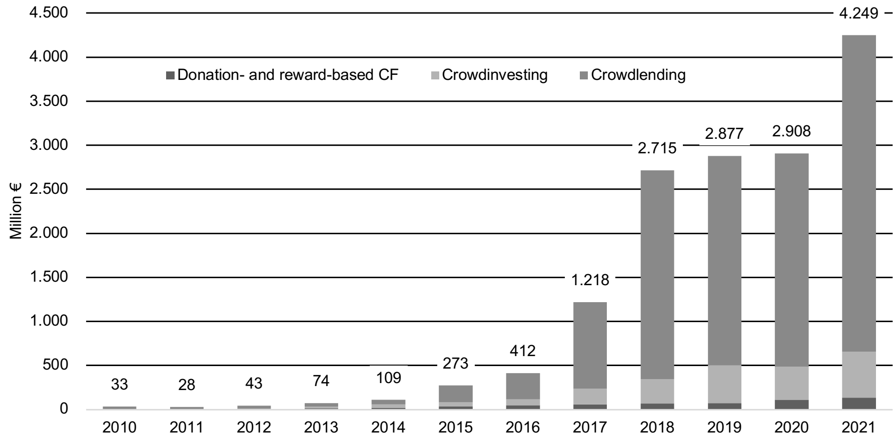


Figure 5.8: Market volume of the segment crowdfunding over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2020 and 2021)

appears to be growing in popularity after being an already large market in which FinTechs provide low entry barriers and funding due to digitization and can take market shares from traditional factoring service providers. The subsegment clearly distinguishes FinTechs from alternative distribution channels of traditional financial intermediaries. If a FinTech is acquired by a bank or no longer operates under its own name, it becomes inactive in our sample. However, we cannot completely rule out the possibility that the FinTech only offers a platform and forwards the volume to traditional financial intermediaries in the background. The largest players on the German market for factoring is *CRX Market* and for credits is *Smava*, followed by *Compeon*, *Aifinyo Factoring* and *Aifinyo Finetrading*.

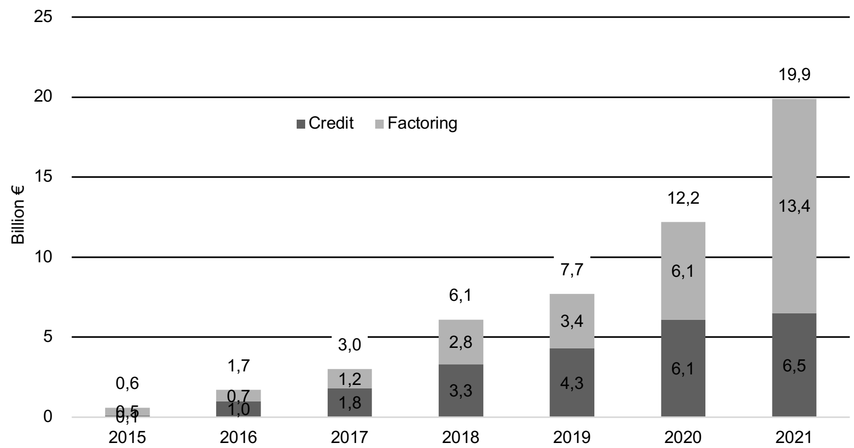


Figure 5.9: Market volume of the subsegment credit and factoring over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2021)

In the investment and banking subsegment, FinTechs focus on traditional banking services such as checking accounts, but typically with more user-friendly functionalities and without cost-driving branch networks. Figure 5.10 shows a linear growth trend over the years reaching a maximum volume of 49.917 million in the year 2021. The largest FinTechs in the subsegment are *Raisin* (in Germany *Weltsparen*), *Deposit Solutions*, *Flaxtex*, *N26* and *Fidor Bank*.

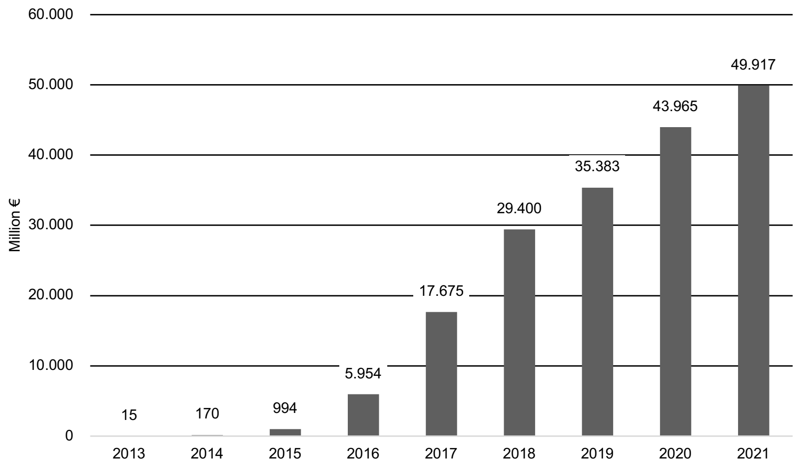


Figure 5.10: Market volume subsegment investment and banking over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2021)

Social trading is a combination of features of online brokers and social networks where a user can follow the trading strategy of another user, which goes so far that the trades can be automatically copied (Glaser and Risius, 2018). The investment strategies use different instruments, such as stocks, exchange-traded funds (ETFs), contracts for difference (CFDs), forex, commodities or cryptocurrencies, depending on the platform. As Figure 5.11 shows, the subsegment of social trading has shown great growth dynamics in recent years. This could be due to the increasing popularity of equity investments in the stock market during the Covid-19 pandemic, as similar dynamics can also be observed in the subsegment robo advice (see Figure 5.12). The market leader on the German market is the Austrian platform *Wikifolio* with a market share of around 75 per cent driving growth and volume in this subsegment, followed by *eToro* and *NagaTrader*.

FinTechs which offer digital and increasingly automated asset management via a platform are assigned to the robo advice subsegment. The personal investment preferences and risk appetite of the investors are taken into account by an algorithm, which allocates the invested capital accordingly. By using robo advisors, investors can achieve diversification effects mostly accompanied by lower volatility and higher returns (D'Acunto et al., 2019). Particularly in the social trading subsegment, we observe the trend towards sustainable investment strategies following the current societal discourse for many robo advice providers. However, one should note that robo advice is a service that even traditional banks are increasingly offering in their online banking, through or without cooperations with FinTechs. As Figure 5.12 shows the assets of German customers managed by robo advisors totaled EUR 10.2 billion at the end of 2021. The German market leader is *Scalable Capital*, followed

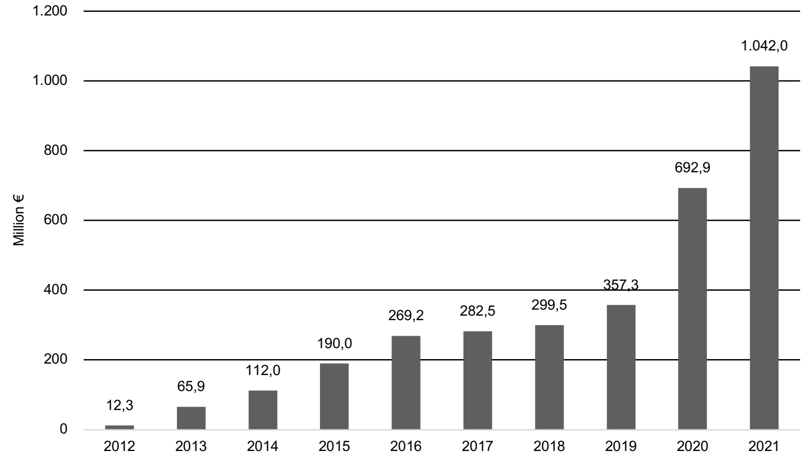


Figure 5.11: Market volume subsegment social trading over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2021)

by *Liqid*, *Quirion* and *Ginmon*.

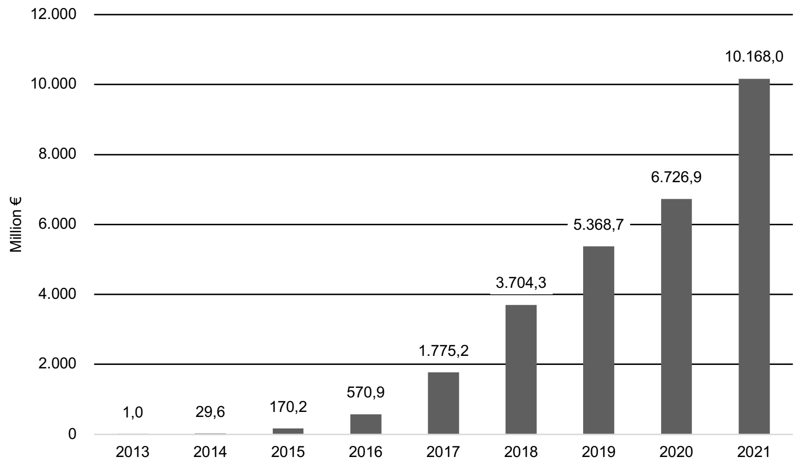


Figure 5.12: Market volume subsegment robo advice over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2021)

To conclude the volume estimates for the year 2021 and the application case of the German FinTech market, we display in Figure 5.13 the sum of the total market volume of the segments financing and asset management over time. We find a steady, linear growth over the years reaching a maximum of 85.3 billion in 2021 in combination with a growth rate of 28 per cent throughout the year 2021. We expect the German FinTech market to establish its position in the market and to further grow. However, the boundaries or demarcation from the traditional banking sector are becoming increasingly blurred in some subsegments due to cooperations or even incorporations with banks.

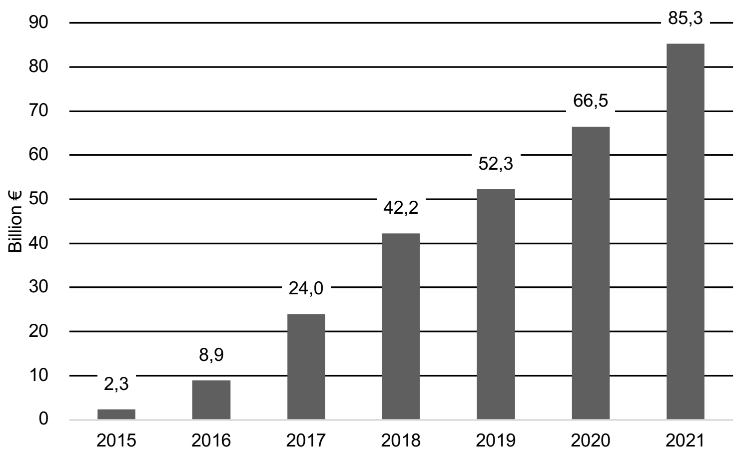


Figure 5.13: Total market volume of the segments financing and asset management over time (Source: Dorfleitner and Hornuf (2023), own calculations for 2020 and 2021)

5.4 Conclusion

The dataset presented is suited to perform descriptive analyses to fully comprehend the complete FinTech market in Germany since its emergence. Especially, the dataset is optimal to obtain a historic perspective. Furthermore, the dataset is useful for everybody interested in the dynamic field of financial technology. Therefore, supervisory authorities, academics as well as practitioners, who need an overview, can benefit from the dataset. Moreover, the nature of the dataset enables researchers to perform further cross-sectional analyses. It provides the possibility of longitudinal analyses of the complete market in Germany to observe trends as well as the maturity of this industry sector.

The entries contain further information that can be used for research that is not necessarily only limited to the market in Germany, but related to the entire international FinTech market. Possible concrete research applications are e.g., founder characteristics in network analysis, the origin of the company to account for the geography of start-ups, the operating status as a success indicator as well as for survival analysis.

Additionally, as demonstrated for the year 2021 the total market volumes of particular FinTech segments can be estimated based on the data. While the evidence on the market volumes presented in this report rather was a quick (and necessarily somewhat imprecise) estimate, the next volume investigations should again be based on the whole cross section of FinTechs in Germany. This is a feasible and rewarding (but laborious) task, which due to the freely accessible data set now can be performed by everyone interested in the German FinTech market.

Acknowledgments

The authors gratefully acknowledge the financial support by the Deutsche Bundesbank. Furthermore, we would like to thank Lars Hornuf, Matthias Schmitt, Lena Wannenmacher and Martina Weber for previous collaboration on this research topic.

Chapter 6

Conclusion

This dissertation contributes to the academic literature on digital finance and financial technology in various ways. The four research papers comprise several distinct analyses of the FinTech sector both at the macro level of the overall FinTech ecosystem and at the micro level examining specific operating sub-segments.

In the first research paper, we study how the General Data Protection Regulation (GDPR) has affected the privacy practices of FinTechs operating in Germany in a pre-and post-GDPR setting. The application of text analysis methods to FinTech's privacy statements suggests that readability has decreased, the texts have become longer, and more standardized language is used, resulting in worse user comprehension. These findings raise the question of whether the GDPR has achieved its goal of protecting natural persons regarding the transparent processing of their personal data. We link the content of the privacy statements to company-and industry-specific determinants and find that before the GDPR became binding, more external investors and a higher legal capital were associated with a higher quantity of data processed and more transparency, but not after that.

The second article examines STO market outcomes in the primary and secondary markets for security tokens and how signaling impacts the behavior of market participants. The conduct of a pre-sale phase before the main funding and the announcement of token transferability are positively related to the funding success and serve as positive quality signals to investors. There is hardly any underpricing on the secondary market, and it is associated with the crypto market sentiment as an external signal. Over various short-term time horizons, security tokens generate both extremely positive and negative returns, which demonstrates the high risk for which investors are not always compensated. The liquidity is lower than on equity markets and utility tokens, especially for decentralized marketplaces with a new model for liquidity provision. Particularly the secondary market for security tokens market needs more professionalism in the valuation and selection of assets.

In the third study, we provide the first empirical insights into tokenized real estate. An extensive blockchain transaction analysis reveals that these tokens enable broad real estate ownership for many retail investors through digital fractional ownership and low entry barriers. Investors do not yet hold well-diversified real estate token portfolios. We find that property-specific fundamentals, crypto market-specific transaction costs, and location-related factors explain most of the success of an STO. For aggregated daily capital flows, we document that investors similarly consider the crypto market sentiment and transaction costs when purchasing tokens; only the transaction costs are relevant for sales. Interestingly, macroeconomic factors have a minor role in capital flows during our observation period.

The last research paper provides an overview of the overall German FinTech market. Furthermore, it offers a valuable application case for the data with market volume estimates for the financing and asset management segment, covering the period until December 2021. We estimate that the total market volume in these segments amounts to €85.3 billion for the year 2021.

Besides its contributions to the academic literature, this dissertation elucidates various implications for stakeholders, including investors, companies, and regulators.

FinTechs enable the democratization of finance for retail investors (Bollaert et al., 2021). Notably, digital tokens present the potential to open up new alternative asset classes, such as real estate, as discussed in Chapter 4, thus, enabling broader diversification and a liquid environment for divestment (refer to Chapter 3). Nonetheless, the findings of this dissertation reveal that investors do not hold well-diversified real estate token portfolios, and secondary markets have not fully matured into liquid marketplaces. Moreover, the results of this dissertation guide investors to discern between high-and low-quality STOs as a basis for their investment decisions. Lastly, investors should be aware that while tokens offer possibilities for broader portfolio diversification, they also expose themselves to inherent technological risks associated with investing in blockchain-based crypto assets.

For businesses, blockchain presents a novel way to gain access to finance, particularly for firms previously unaccounted for by traditional financial institutions (Erel and Liebersohn, 2022) and to illustrate the successful disintermediation of traditional players such as mortgage lenders and security custodians. Given the nascent market structures and sometimes high information asymmetries, the findings of this dissertation can assist entrepreneurs in successfully designing their offerings to transmit positive quality signals to prospective investors. Companies would be well advised to include a pre-sale phase before the main offering to gather pricing-relevant information and declare their intentions to make tokens tradable on the secondary market. Simultaneously, young companies in the company-building process must not disregard compliance with data protection regulations and transparent communication with their users to avoid fines and potentially confer a competitive advantage over rivals.

Given the increasing prevalence of data-driven business models and the rise of artificial intelligence, the importance of privacy regulation is anticipated to amplify in the future, particularly for FinTechs. These developments are countered by the GDPR, which has been adopted for some time. Policymakers can ascertain the impact and potential unintended consequences of the GDPR with the findings derived from this dissertation to undertake possible adjustments and offer clear guidance for privacy statement design in the future.

In general, the primary objective of FinTech regulation is to establish a framework that promotes innovation to enhance financial inclusion and economic growth while safeguarding investors or users and the financial system and preserving overall financial stability (Allen et al., 2021). The empirical evidence presented in this dissertation underscores that investors are substantially affected by market sentiment and trends that do not fully reflect asset fundamentals. This finding highlights the need for consumer protection and specific crypto regulation. As part of the Digital Finance Package, the Council of the European Union adopted the Markets in Crypto-assets Regulation (MiCA) in May 2023, thereby establishing the first comprehensive legal framework for the crypto industry at the European level. Security tokens usually fall under the purview of securities regulation, and MiCA confirms that and provides a legal framework for crypto assets. Consequently, MiCA assures legal certainty to all crypto industry stakeholders in the EU. Additionally, it creates a potential competitive advantage over the uncertain case law system in the US. This new regulation is likely to promote the future advancement of the crypto industry and, thus, bolster the market for digital assets in Europe.

This dissertation has some limitations that point toward avenues for future research. In the study on the impact of the GDPR on FinTechs, we assume that the privacy practices of FinTech firms are consistent with what they state in their privacy statements, the supply side of privacy (Ramadorai et al., 2021). Clearly, users have no choice but to accept the terms and conditions if they want to use a company's service or product (Aridor et al., 2020). Consequently, one avenue for future research is to compare what companies claim in their privacy statements and what they actually do regarding privacy practices.

Given the relatively recent and dynamic nature of FinTech, empirical research encounters constraints concerning several aspects of data availability, as outlined in the Introduction. Consequently, the generalizability of some findings presented in this dissertation may be limited. Specifically, the analysis of macroeconomic factors on real estate token capital flows in Chapter 4 yields inconclusive and insignificant effects, potentially attributable to the observation period ending in December 2021. Throughout this period, treasury rates remained relatively constant. However, since 2022, the economy has experienced rising inflation rates leading to an interest rate increase. Moreover, the crypto industry experienced various incidents in 2022 that substantially impacted overall confidence. As this dissertation's findings emphasize the importance of sentiment on the crypto market, it is essential to consider these developments in future analyses. More generally speaking, certain findings in this dissertation pertain to specific observation periods, countries, jurisdictions, or asset classes. This issue will resolve itself with increasing familiarity and acceptance of blockchain technology in the population and a broader data basis over time. As such, future research suggests testing and verifying these results within a more extensive sample, ideally in a setting that allows providing causal evidence.

FinTech firms operate within a highly dynamic environment that disrupts traditional market structures in the financial industry, marked by the tension between technological innovation and changing regulatory frameworks. As a result, digitization in the financial sector remains a constantly evolving field that offers many opportunities for future academic research across a broad spectrum of topics.

Bibliography

Acquisti, A., 2004. Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the 5th ACM Conference on Electronic Commerce. EC '04. Association for Computing Machinery, New York, pp. 21–29.

Acquisti, A., Brandimarte, L., Loewenstein, G., 2015. Privacy and human behavior in the age of information. Science 347 (6221), 509–514.

Acquisti, A., Brandimarte, L., Loewenstein, G., 2020. Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. Journal of Consumer Psychology 30 (4), 736–758.

Acquisti, A., Taylor, C., Wagman, L., 2016. The economics of privacy. Journal of Economic Literature 54 (2), 442–492.

Adhami, S., Giudici, G., Martinazzi, S., 2018. Why do businesses go crypto? An empirical analysis of initial coin offerings. Journal of Economics and Business 100, 64–75.

Admati, A. R., Pfleiderer, P., Zechner, J., 1994. Large shareholder activism, risk sharing, and financial market equilibrium. Journal of Political Economy 102 (6), 1097–1130.

Agarwal, S., Steyskal, S., Antunovic, F., Kirrane, S., 2018. Legislative compliance assessment: Framework, model and GDPR instantiation. In: Medina, M., Mitrakas, A., Rannenberg, K., Schweighofer, E., Tsouroulas, N. (Eds.), Privacy Technologies and Policy. Springer International Publishing, Cham, pp. 131–149.

Aggarwal, R. K., Krigman, L., Womack, K. L., 2002. Strategic IPO underpricing, information momentum, and lockup expiration selling. Journal of Financial Economics 66 (1), 105–137.

Ahlers, G. K. C., Cumming, D., Günther, C., Schweizer, D., 2015. Signaling in equity crowdfunding. Entrepreneurship Theory and Practice 39 (4), 955–980.

Ajaz, T., Kumar, A. S., 2018. Herding in crypto-currency markets. Annals of Financial Economics 13 (2), 1850006.

Allen, F., Xian, G., Jagtiani, J., 2021. A survey of Fintech research and policy discussion. Review of Corporate Finance 1 (3–4), 259–339.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5 (1), 31–56.

Amihud, Y., Mendelson, H., Pedersen, L. H., 2006. Liquidity and asset prices. Foundations and Trends in Finance 1 (4), 269–364.

Amsden, R., Schweizer, D., 2019. Are blockchain crowdsales the new 'gold rush'? Success determinants of initial coin offerings. Working paper 3163849, Social Science Research Network.

An, J., Duan, T., Hou, W., Xu, X., 2019. Initial coin offerings and entrepreneurial finance: The role of founders characteristics. The Journal of Alternative Investments 21 (4), 26–40.

Ante, L., Fiedler, I., 2020. Cheap signals in security token offerings (STOs). Quantitative Finance and Economics 4 (4), 608–639.

Aoyagi, J., 2020. Liquidity provision by automated market makers. Working paper 3674178, Social Science Research Network.

Aridor, G., Che, Y.-K., Salz, T., 2020. The economic consequences of data privacy regulation: Empirical evidence from GDPR. Working Paper 26900, National Bureau of Economic Research.

Arora, C., Sabetzadeh, M., Briand, L. C., Zimmer, F., 2014. Requirement boilerplates: Transition from manually-enforced to automatically-verifiable natural language patterns. In: 2014 IEEE 4th International Workshop on Requirements Patterns (RePa). pp. 1–8.

Article 29 Working Party, 2018. Guidelines on Transparency under Regulation 2016/679. WP260 rev.01.

Aruoba, S. B., Diebold, F. X., Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business & Economic Statistics 27 (4), 417–427.

Aspris, A., Foley, S., Svec, J., Wang, L., 2021. Decentralized exchanges: The "wild west" of cryptocurrency trading. International Review of Financial Analysis 77, 101845.

Auer, R., Tercero-Lucas, D., 2022. Distrust or speculation? The socioeconomic drivers of U.S. cryptocurrency investments. Journal of Financial Stability 62, 101066.

BaFin, 2023. Crypto tokens. https://www.bafin.de/dok/18463500, accessed 15 May 2023.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61 (4), 1645–1680.

Bakos, Y., Marotta-Wurgler, F., Trossen, D. R., 2014. Does anyone read the fine print? Consumer attention to standard-form contracts. The Journal of Legal Studies 43 (1), 1–35.

Bamberger, R., Vanecek, E., 1984. Lesen-Verstehen-Lernen-Schreiben: Die Schwierigkeitsstufen von Texten in deutscher Sprache. Jugend und Volk Verlagsgesellschaft, Vienna.

Barbon, A., Ranaldo, A., 2022. On the quality of cryptocurrency markets: Centralized versus decentralized exchanges. Working paper, University of St. Gallen, School of Finance Research Paper Forthcoming, Swiss Finance Institute Research Paper No. 22-38.

Baum, A., 2021. Tokenization–the future of real estate investment? The Journal of Portfolio Management 48 (1).

Baum, J. A., Silverman, B. S., 2004. Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. Journal of Business Venturing 19 (3), 411–436.

Becher, S. I., Benoliel, U., 2021. Law in books and law in action: The readability of privacy policies and the GDPR. In: Mathis, K., Avishalom, T. (Eds.), Consumer Law & Economics: Economic Analysis of Law in European Legal Scholarship. Vol. 9. Springer, New York, pp. 179–204.

Belleflamme, P., Lambert, T., Schwienbacher, A., 2014. Crowdfunding: Tapping the right crowd. Journal of Business Venturing 29 (5), 585–609.

Benedetti, H., Kostovetsky, L., 2021. Digital Tulips? Returns to investors in initial coin offerings. Journal of Corporate Finance 66, 101786.

Benedetti, H., Rodríguez-Garnica, G., 2023. Tokenized assets and securities. In: Baker, H., Benedetti, H., Nikbakht, E., Smith, S. (Eds.), The Emerald Handbook on Cryptoassets: Investment Opportunities and Challenges. Emerald Publishing Limited, Bingley, pp. 107–121.

Berg, T., Burg, V., Gombović, A., Puri, M., 2020. On the rise of FinTechs: Credit scoring using digital footprints. The Review of Financial Studies 33 (7), 2845–2897.

Bergh, D. D., Connelly, B. L., Ketchen Jr, D. J., Shannon, L. M., 2014. Signalling theory and equilibrium in strategic management research: An assessment and a research agenda. Journal of Management Studies 51 (8), 1334–1360.

Bernstein, S., Korteweg, A., Laws, K., 2017. Attracting early-stage investors: Evidence from a randomized field experiment. The Journal of Finance 72 (2), 509–538.

Betzing, J. H., Tietz, M., vom Brocke, J., Becker, J., 2020. The impact of transparency on mobile privacy decision making. Electronic Markets 30, 607–625.

Biais, B., Bisiere, C., Bouvard, M., Casamatta, C., Menkveld, A. J., 2023. Equilibrium bitcoin pricing. The Journal of Finance 78 (2), 967–1014.

Bianchi, D., 2020. Cryptocurrencies as an asset class? An empirical assessment. The Journal of Alternative Investments 23 (2), 162–179.

Bianchi, D., Babiak, M., 2022. A risk-based explanation of cryptocurrency returns. Working Paper 3935934, Social Science Research Network.

Biasiotti, M., Francesconi, E., Palmirani, M., Sartor, G., Vitali, F., 2008. Legal informatics and management of legislative documents. Working Paper 2, Global Center for ICT in Parliament.

Block, J., Hornuf, L., Moritz, A., 2018. Which updates during an equity crowdfunding campaign increase crowd participation? Small Business Economics 50, 3–27.

Block, J. H., Groh, A., Hornuf, L., Vanacker, T., Vismara, S., 2020. The entrepreneurial finance markets of the future: a comparison of crowdfunding and initial coin offerings. Small Business Economics 57, 865–882.

Bollaert, H., de Silanes, F. L., Schwienbacher, A., 2021. Fintech and access to finance. Journal of Corporate Finance 68, 101941.

Booth, J. R., Smith, R. L., 1986. Capital raising, underwriting and the certification hypothesis. Journal of Financial Economics 15 (1), 261–281.

Boreiko, D., Risteski, D., 2021. Serial and large investors in initial coin offerings. Small Business Economics 57, 1053–1071.

Bourassa, S. C., Haurin, D. R., Haurin, J. L., Hoesli, M., Sun, J., 2009. House price changes and idiosyncratic risk: The impact of property characteristics. Real Estate Economics 37 (2), 259–278.

Bouri, E., Gupta, R., Roubaud, D., 2019. Herding behaviour in cryptocurrencies. Finance Research Letters 29, 216–221.

Brau, J. C., Fawcett, S. E., 2006. Initial public offerings: An analysis of theory and practice. The Journal of Finance 61 (1), 399–436.

Brauneis, A., Mestel, R., Riordan, R., Theissen, E., 2021. How to measure the liquidity of cryptocurrency markets? Journal of Banking & Finance 124, 106041.

Brown, S. V., Tucker, J. W., 2011. Large-sample evidence on firms' year-over-year MD&A modifications. Journal of Accounting Research 49 (2), 309–346.

Buterin, V., 2013. Ethereum: A next-generation smart contract and decentralized application platform. White paper.

Caporin, M., Costola, M., Garibal, J.-C., Maillet, B., 2022. Systemic risk and severe economic downturns: A targeted and sparse analysis. Journal of Banking & Finance 134, 106339.

Carhart, M. M., 1997. On Persistence in Mutual Fund Performance. The Journal of Finance 52 (1), 57–82.

Chang, C., 2020. From securitization to tokenization. In: Pentland, A., Lipton, A., Hardjono, T. (Eds.), Building the New Economy. MIT Press.

Chang, E. C., Cheng, J. W., Khorana, A., 2000. An examination of herd behavior in equity markets: An international perspective. Journal of Banking & Finance 24 (10), 1651–1679.

Chanson, M., Gjoen, J., Risius, M., Wortmann, F., 2018. Initial coin offerings (ICOs): The role of social media for organizational legitimacy and underpricing. In: 39th International Conference on Information Systems (ICIS). San Francisco.

Chemla, G., Tinn, K., 2020. Learning through crowdfunding. Management Science 66 (5), 1783–1801.

Chen, M. A., Wu, Q., Yang, B., 2019. How valuable is FinTech innovation? The Review of Financial Studies 32 (5), 2062–2106.

Chen, Y., Pereira, I., Patel, P. C., 2021. Decentralized governance of digital platforms. Journal of Management 47 (5), 1305–1337.

Chiang, T. C., Zheng, D., 2010. An empirical analysis of herd behavior in global stock markets. Journal of Banking & Finance 34 (8), 1911–1921.

Chod, J., Trichakis, N., Yang, S. A., 2022. Platform tokenization: Financing, governance, and moral hazard. Management Science 68 (9), 6411–6433.

Cohen, J. E., 2000. Examined lives: Informational privacy and the subject as object. Stanford Law Review 52 (5), 1373–1438.

Cohen, L., Malloy, C., Nguyen, Q., 2020. Lazy prices. The Journal of Finance 75 (3), 1371–1415.

Colombo, M. G., Franzoni, C., Rossi-Lamastra, C., 2015. Internal social capital and the attraction of early contributions in crowdfunding. Entrepreneurship Theory and Practice 39 (1), 75–100.

Cong, L. W., He, Z., 2019. Blockchain disruption and smart contracts. The Review of Financial Studies 32 (5), 1754–1797.

Cong, L. W., Li, B., Zhang, Q. T., 2021. Alternative data in FinTech and business intelligence. In: Pompella, M., Matousek, R. (Eds.), The Palgrave Handbook of FinTech and Blockchain. Palgrave Macmillan, Cham, pp. 217–242.

Cong, L. W., Li, Y., Wang, N., 2022. Token-based platform finance. Journal of Financial Economics 144 (3), 972–991.

Connelly, B. L., Certo, S. T., Ireland, R. D., Reutzel, C. R., 2011. Signaling theory: A review and assessment. Journal of Management 37 (1), 39–67.

Contissa, G., Docter, K., Lagioia, F., Lippi, M., Micklitz, H.-W., Palka, P., Sartor, G., Torroni, P., 2018. Claudette meets GDPR: Automating the evaluation of privacy policies using artificial intelligence. Working Paper 3208596, Social Science Research Network.

Corwin, S. A., Schultz, P. H., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. The Journal of Finance 67 (2), 719–760.

Courtney, C., Dutta, S., Li, Y., 2017. Resolving information asymmetry: Signaling, endorsement, and crowdfunding success. Entrepreneurship Theory and Practice 41 (2), 265–290.

Cronqvist, H., Hogfeldt, P., Nilsson, M., 2001. Why agency costs explain diversification discounts. Real Estate Economics 29 (1), 85–126.

Cudd, M., Davis, H. E., Eduardo, M., 2006. Mimicking behavior in repurchase decisions. Journal of Behavioral Finance 7 (4), 222–229.

Cumming, D., Meoli, M., Vismara, S., 2019. Investors' choices between cash and voting rights: Evidence from dual-class equity crowdfunding. Research Policy 48 (8), 103740.

Cumming, D. J., Leboeuf, G., Schwienbacher, A., 2020. Crowdfunding models: Keep-it-all vs. all-or-nothing. Financial Management 49 (2), 331–360.

Cumming, D. J., Schwienbacher, A., 2018. Fintech venture capital. Corporate Governance: An International Review 26 (5), 374–389.

Da, Z., Engelberg, J., Gao, P., 10 2014. The sum of all FEARS investor sentiment and asset prices. The Review of Financial Studies 28 (1), 1–32.

D'Acunto, F., Prabhala, N., Rossi, A. G., 2019. The promises and pitfalls of robo-advising. The Review of Financial Studies 32 (5), 1983–2020.

De Clercq, D., Dimov, D., 2008. Internal knowledge development and external knowledge access in venture capital investment performance. Journal of Management Studies 45 (3), 585–612.

De Crescenzo, V., Ribeiro-Soriano, D. E., Covin, J. G., 2020. Exploring the viability of equity crowdfunding as a fundraising instrument: A configurational analysis of contingency factors that lead to crowdfunding success and failure. Journal of Business Research 115, 348–356.

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., Holz, T., 2019. We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. In: 26th Annual Network and Distributed System Security Symposium, NDSS 2019. The Internet Society, San Diego.

Demir, A., Pesqué-Cela, V., Altunbas, Y., Murinde, V., 2022. Fintech, financial inclusion and income inequality: a quantile regression approach. The European Journal of Finance 28 (1), 86–107.

Deutsche Bundesbank, 2023. Fintech–technology-enabled financial innovation. https://www.bundesbank.de/en/tasks/banking-supervision/individual-aspects/fintechs/fintech-technology-enabled-financial-innovation-622840, accessed 2 May 2023.

Diamond, D. W., 1984. Financial intermediation and delegated monitoring. The Review of Economic Studies 51 (3), 393–414.

Dinev, T., Hart, P., 2006. An extended privacy calculus model for e-commerce transactions. Information Systems Research 17 (1), 61–80.

Ding, X. S., Ni, Y., Zhong, L., 2016. Free float and market liquidity around the world. Journal of Empirical Finance 38, 236–257.

Dobkin, A., 2018. Information fiduciaries in practice: Data privacy and user expectations. Berkeley Technology Law Journal 33 (1), 1–52.

Dorfleitner, G., Hornuf, L., 2019. FinTech and Data Privacy in Germany: An Empirical Analysis with Policy Recommendations. Springer International Publishing, Cham.

Dorfleitner, G., Hornuf, L., 2023. Der deutsche FinTech-Markt. In: Voß, T. (Ed.), Recht der FinTechs. De Gruyter, Berlin.

Dorfleitner, G., Hornuf, L., Kreppmeier, J., 2023. Promise not fulfilled: FinTech, data privacy, and the GDPR. Electronic Markets, forthcoming.

Dorfleitner, G., Hornuf, L., Schmitt, M., Weber, M., 2017. FinTech in Germany. Springer International Publishing, Cham.

Dorfleitner, G., Hornuf, L., Wannenmacher, L., 2020. Der deutsche FinTech-Markt im Jahr 2020. ifo Schnelldienst 73, 33–40.

Dorfleitner, G., Kreppmeier, J., Laschinger, R., 2022. Dataset of German FinTech companies: A market overview. Mendeley Data V2, https://doi.org/10.17632/438ytjyzxk.2.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., Kammler, J., 2016. Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms. Journal of Banking & Finance 64, 169–187.

Drasch, B. J., Schweizer, A., Urbach, N., 2018. Integrating the 'troublemakers': A taxonomy for cooperation between banks and fintechs. Journal of Economics and Business 100, 26–42.

Drobetz, W., Momtaz, P. P., Schröder, H., 2019. Investor sentiment and initial coin offerings. The Journal of Alternative Investments 21 (4), 41–55.

Duchesneau, D. A., Gartner, W. B., 1990. A profile of new venture success and failure in an emerging industry. Journal of Business Venturing 5 (5), 297–312.

Earp, J. B., Anton, A. I., Aiman-Smith, L., Stufflebeam, W. H., 2005. Examining internet privacy policies within the context of user privacy values. IEEE Transactions on Engineering Management 52 (2), 227–237.

Engert, A., Hornuf, L., 2018. Market standards in financial contracting: The Euro's effect on debt securities. Journal of International Money and Finance 85, 145–162.

Erel, I., Liebersohn, J., 2022. Can fintech reduce disparities in access to finance? Evidence from the paycheck protection program. Journal of Financial Economics 146 (1), 90–118.

Ermakova, T., Baumann, A., Fabian, B., Krasnova, H., 2014. Privacy policies and users' trust: Does readability matter? In: Americas Conference on Information Systems. Savannah.

Esrig, D., Hudgins, M. C., Cerreta, L., 2011. Revisiting the impact of large assets on real estate portfolio returns. The Journal of Portfolio Management 37 (5), 125–136.

Ethereum.org, 2022. Gas and fees. https://ethereum.org/en/developers/docs/gas/, accessed 28 November 2022.

European Banking Authority, 2017. Discussion paper on the EBA's approach to financial technology (FinTech). Discussion Paper EBA/DP/2017/02.

European Commission, 2020. Digital finance strategy for the EU. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2020) 240.

Fabian, B., Ermakova, T., Lentz, T., 2017. Large-scale readability analysis of privacy policies. In: Proceedings of the International Conference on Web Intelligence. WI '17. Association for Computing Machinery, New York, pp. 18–25.

Fahlenbrach, R., Frattaroli, M., 2021. ICO investors. Financial Markets and Portfolio Management 35, 1–59.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Feldman, M., Kenney, M., Lissoni, F., 2015. The new data frontier: Special issue of research policy. Research Policy 44 (9), 1629–1632.

Felipe, I. J. S., Mendes-Da-Silva, W., Leal, C. C., Braun Santos, D., 2022. Reward crowdfunding campaigns: Time-to-success analysis. Journal of Business Research 138, 214–228.

Felix, T. H., von Eije, H., 2019. Underpricing in the cryptocurrency world: Evidence from initial coin offerings. Managerial Finance 45 (4), 563–578.

Fernback, J., Papacharissi, Z., 2007. Online privacy as legal safeguard: The relationship among consumer, online portal, and privacy policies. New Media & Society 9 (5), 715–734.

Financial Stability Board, 2017. Financial stability implications from FinTech: Supervisory and regulatory issues that merit authorities' attention. Reports to the G20.

Firtel, K. B., 1999. Plain English: A reappraisal of the intended audience of disclosure under the securities act of 1933. Southern California Law Review 72, 851–898.

Fisch, C., 2019. Initial coin offerings (ICOs) to finance new ventures. Journal of Business Venturing 34 (1), 1–22.

Fisch, C., Masiak, C., Vismara, S., Block, J., 2021. Motives and profiles of ICO investors. Journal of Business Research 125, 564–576.

Fisch, C., Meoli, M., Vismara, S., 2022. Does blockchain technology democratize entrepreneurial finance? An empirical comparison of ICOs, venture capital, and REITs. Economics of Innovation and New Technology 31 (1–2), 70–89.

Fisch, C., Momtaz, P. P., 2020. Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). Journal of Corporate Finance 64, 101679.

Flesch, R., 1948. A new readability yardstick. Journal of Applied Psychology 32 (3), 221.

Florysiak, D., Schandlbauer, A., 2022. Experts or charlatans? ICO analysts and white paper informativeness. Journal of Banking & Finance 139, 106476.

Franzke, S. A., 2004. Underpricing of venture-backed and non venture-backed IPOs: Germany's Neuer Markt. In: Giudici, G., Roosenboom, P. (Eds.), The Rise and Fall of Europe's New Stock Markets. Vol. 10 of Advances in Financial Economics. Emerald Group Publishing Limited, Bingley, pp. 201–230.

Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., Zbinden, P., 2020. BigTech and the changing structure of financial intermediation. Economic Policy 34 (100), 761–799.

Gai, K., Qiu, M., Sun, X., Zhao, H., 2017. Security and privacy issues: A survey on fintech. In: Qiu, M. (Ed.), Smart Computing and Communication. Springer International Publishing, Cham, pp. 236–247.

Gan, J. R., Tsoukalas, G., Netessine, S., 2021. Initial coin offerings, speculation, and asset tokenization. Management Science 67 (2), 914–931.

Gao, Q., Lin, M., Sias, R. W., 2023. Words matter: The role of texts in online credit markets. Journal of Financial and Quantitative Analysis 58 (1), 1–28.

Gazel, M., Schwienbacher, A., 2021. Entrepreneurial fintech clusters. Small Business Economics 57, 883–903.

Geltner, D., Miller, N. G., Clayton, J., Eichholtz, P., 2014. Commercial real estate, 3rd Edition. OnCourse Learning, Mason OH.

Giudici, G., Roosenboom, P., 2004. Pricing initial public offerings on Europe's New Stock Markets. In: Giudici, G., Roosenboom, P. (Eds.), The Rise and Fall of Europe's New Stock Markets. Vol. 10 of Advances in Financial Economics. Emerald Group Publishing Limited, Bingley, pp. 25–59.

Glaser, F., Risius, M., 2018. Effects of transparency: Analyzing social biases on trader performance in social trading. Journal of Information Technology 33 (1), 19–30.

Goetzmann, W. N., Kumar, A., 2008. Equity portfolio diversification. Review of Finance 12 (3), 433–463.

Goldberg, S. G., Johnson, G. A., Shriver, S. K., 2021. Regulating privacy online: An economic evaluation of the GDPR. Working Paper 3421731, Social Science Research Network.

Goldstein, I., Jiang, W., Karolyi, G. A., 2019. To FinTech and beyond. The Review of Financial Studies 32 (5), 1647–1661.

Gomber, P., Kauffman, R. J., Parker, C., Weber, B. W., 2018. On the Fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. Journal of Management Information Systems 35 (1), 220–265.

Gomber, P., Koch, J.-A., Siering, M., 2017. Digital finance and FinTech: Current research and future research directions. Journal of Business Economics 87 (5), 537–580.

Green, T. C., Hwang, B.-H., 2012. Initial public offerings as lotteries: Skewness preference and first-day returns. Management Science 58 (2), 432–444.

Gryglewicz, S., Mayer, S., Morellec, E., 2021. Optimal financing with tokens. Journal of Financial Economics 142 (3), 1038–1067.

Gunning, R., 1952. The Technique of Clear Writing. McGraw-Hill, New York.

Guo, Y., Liang, C., 2016. Blockchain application and outlook in the banking industry. Financial Innovation 2 (1), 24.

Gupta, A., Rathod, J., Patel, D., Bothra, J., Shanbhag, S., Bhalerao, T., 2020. Tokenization of real estate using blockchain technology. In: Zhou, J., Conti, M., Ahmed, C. M., Au, M. H., Batina, L., Li, Z., Lin, J., Losiouk, E., Luo, B., Majumdar, S., Meng, W., Ochoa, M., Picek, S., Portokalidis, G., Wang, C., Zhang, K. (Eds.), Applied Cryptography and Network Security Workshops. Springer International Publishing, Cham, pp. 77–90.

Hackober, C., Bock, C., 2021. Which investors' characteristics are beneficial for initial coin offerings? Evidence from blockchain technology-based firms. Journal of Business Economics 91, 1085–1124.

Haddad, C., Hornuf, L., 2019. The emergence of the global fintech market: Economic and technological determinants. Small Business Economics 53 (1), 81–105.

Hainz, C., Hornuf, L., Klöhn, L., Brauer, B., Ehrenfried, F., Engelmann, G., 2017. Die Befreiungsvorschriften des Kleinanlegerschutzgesetzes. ifo Schnelldienst 70 (6), 26–35.

Hainz, C., Hornuf, L., Nagel, L., Reiter, S., Stenzhorn, E., 2019. Die Befreiungsvorschriften des Kleinanlegerschutzgesetzes: Eine Follow-up Studie. ifo Schnelldienst 72 (9), 26–37.

Hajduk, P., 2021. The powers of the supervisory body in GDPR as a basis for shaping the practices of personal data processing. Review of European and Comparative Law 45 (2), 57–75.

Halvorsen, R., Palmquist, R., 1980. The interpretation of dummy variables in semilogarithmic equations. American Economic Review 70 (3), 474–475.

Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K. G., Aberer, K., 2018. Polisis: Automated analysis and presentation of privacy policies using deep learning. In: USENIX Security Symposium. pp. 531–548.

Hartmann, P. M., Zaki, M., Feldmann, N., Neely, A., 2016. Capturing value from big data - a taxonomy of data-driven business models used by start-up firms. International Journal of Operations & Production Management 36 (10), 1382–1406.

Hartzell, J. C., Sun, L., Titman, S., 2014. Institutional investors as monitors of corporate diversification decisions: Evidence from real estate investment trusts. Journal of Corporate Finance 25, 61–72.

Hasbrouck, J., Rivera, T. J., Saleh, F., 2022. The need for fees at a DEX: How increases in fees can increase DEX trading volume. Working Paper 4192925, Social Science Research Network.

Heckman, J. J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 5 (4).

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.

Hendershott, T., Zhang, X. M., Zhao, J. L., Zheng, Z. E., 2021. FinTech as a game changer: Overview of research frontiers. Information Systems Research 32 (1), 1–17.

Herfindahl, O., 1950. Concentration in the steel industry.

Hillebrand, K., Hornuf, L., Müller, B., Vrankar, D., 2023. The social dilemma of big data: Donating personal data to promote social welfare. Information and Organization 33 (1), 100452.

Hinzen, F. J., John, K., Saleh, F., 2022. Bitcoin's limited adoption problem. Journal of Financial Economics 144 (2), 347–369.

Hirschman, A. O., 1964. The paternity of an index. The American Economic Review 54 (5), 761–762.

Hornuf, L., Kloehn, L., Schilling, T., 2018a. Financial contracting in crowdinvesting: Lessons from the German market. German Law Journal 19 (3), 509–578.

Hornuf, L., Klus, M. F., Lohwasser, T. S., Schwienbacher, A., 2021a. How do banks interact with fintech startups? Small Business Economics 57, 1505–1526.

Hornuf, L., Schilling, T., Schwienbacher, A., 2021b. The relevance of investor rights in crowdinvesting. Journal of Corporate Finance, 101927.

Hornuf, L., Schmitt, M., Stenzhorn, E., 2018b. Equity crowdfunding in Germany and the United Kingdom: Follow-up funding and firm failure. Corporate Governance: An International Review 26 (5), 331–354.

Hornuf, L., Schwienbacher, A., 2018. Market mechanisms and funding dynamics in equity crowdfunding. Journal of Corporate Finance 50, 556–574.

Howell, S. T., Niessner, M., Yermack, D., 2018. Initial coin offerings: Financing growth with cryptocurrency token sales. Working paper 564/2018, European Corporate Governance Institute.

Howell, S. T., Niessner, M., Yermack, D., 2020. Initial coin offerings: Financing growth with cryptocurrency token sales. The Review of Financial Studies 33 (9), 3925–3974.

Hsu, D. H., 2006. Venture capitalists and cooperative start-up commercialization strategy. Management Science 52 (2), 204–219.

Ingram Bogusz, C., 2018. Digital traces, ethics, and insight: Data-driven services in FinTech. In: Teigland, R., Siri, S., Larsson, A., Puertas, A. M., Ingram Bogusz, C. (Eds.), The Rise and Development of Fintech: Accounts of Disruption from Sweden and Beyond. Routledge, London, pp. 207–222.

Jiang, X., Rodríguez Jr, I. M., Zhang, Q., 2023. Macroeconomic fundamentals and cryptocurrency prices: A common trend approach. Financial Management 52 (1), 181–198.

Jiang, Y., Ho, Y.-C. C., Yan, X., Tan, Y., 2020. When online lending meets real estate: Examining investment decisions in lending-based real estate crowdfunding. Information Systems Research 31 (3), 715–730.

Kahan, M., Klausner, M., 1997. Standardization and innovation in corporate contracting (or the economics of boilerplate). Virginia Law Review 83 (4), 713–770.

Kaur, J., Dara, R. A., Obimbo, C., Song, F., Menard, K., 2018. A comprehensive keyword analysis of online privacy policies. Information Security Journal: A Global Perspective 27 (5-6), 260–275.

Kiel, K. A., Zabel, J. E., 2008. Location, location, location: The 3L approach to house price determination. Journal of Housing Economics 17 (2), 175–190.

Kiss, I., Stehle, R., 2002. Underpricing and long-term performance of initial public offerings at Germanys Neuer Markt, 1997-2001. Working paper, Humboldt-University Berlin.

Klus, M. F., Lohwasser, T. S., Holotiuk, F., Moormann, J., 2019. Strategic alliances between banks and fintechs for digital innovation: Motives to collaborate and types of interaction. Journal of Entrepreneurial Finance 21 (1), 1–23.

Konashevych, O., 2020. General concept of real estate tokenization on blockchain: The right to choose. European Property Law Journal 9 (1), 21–66.

Kondra, A. Z., Hinings, C. R., 1998. Organizational diversity and change in institutional theory. Organization Studies 19 (5), 743–767.

KPMG International, 2023. Pulse of Fintech H2'22: Global analysis of fintech investment. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2023/02/pulse-of-fintech-h2-22-web-file.pdf, data provided by PitchBook.

Kubick, T. R., Lynch, D. P., Mayberry, M. A., Omer, T. C., 2015. Product market power and tax avoidance: Market leaders, mimicking strategies, and stock returns. The Accounting Review 90 (2), 675–702.

Lambert, T., Liebau, D., Roosenboom, P., 2022. Security token offerings. Small Business Economics 59, 299–325.

Laursen, K., Salter, A. J., 2014. The paradox of openness: Appropriability, external search and collaboration. Research Policy 43 (4), 867–878.

Lee, I., Shin, Y. J., 2018. Fintech: Ecosystem, business models, investment decisions, and challenges. Business Horizons 61 (1), 35–46.

Lee, J., Li, T., Shin, D., 2022. The wisdom of crowds in FinTech: Evidence from initial coin offerings. The Review of Corporate Finance Studies 11 (1), 1–46.

Lehar, A., Parlour, C. A., 2022. Decentralized exchanges. Working paper, University of Calgary, University of California, Berkeley.

Lewis, S. D., Colvard, R. G., Adams, C. N., 2008. A comparison of the readability of privacy statements of banks, credit counseling companies, and check cashing companies. Journal of Organizational Culture, Communications and Conflict 12 (2), 87–93.

Li, H., Yu, L., He, W., 2019. The impact of GDPR on global technology development. Journal of Global Information Technology Management 22 (1), 1–6.

Li, J., Li, J., Zhu, X., Yao, Y., Casu, B., 2020. Risk spillovers between fintech and traditional financial institutions: Evidence from the US. International Review of Financial Analysis 71, 101544.

Lin, L. X., 2019. Deconstructing decentralized exchanges. Stanford Journal of Blockchain Law & Policy 2 (1), 58–77.

Linden, T., Khandelwal, R., Harkous, H., Fawaz, K., 2020. The privacy policy landscape after the GDPR. In: Proceedings on Privacy Enhancing Technologies. pp. 47–64.

Lindgreen, E. R., 2018. Privacy from an economic perspective. In: The Handbook of Privacy Studies: An Interdisciplinary Introduction. Amsterdam University Press, Amsterdam, pp. 181–208.

Ling, D. C., Archer, W. R., 2021. Real estate principles: A value approach, 6th Edition. McGraw-Hill Education.

Litimi, H., BenSaïda, A., Bouraoui, O., 2016. Herding and excessive risk in the American stock market: A sectoral analysis. Research in International Business and Finance 38, 6–21.

Liu, N., Duncan, R., Chapman, A., 2020a. A critical review of distributed ledger technology and its applications in real estate. RICS research report.

Liu, X., Farahani, B., Firouzi, F., 2020b. Distributed ledger technology. In: Firouzi, F., Chakrabarty, K., Nassif, S. (Eds.), Intelligent Internet of Things. Springer, Cham, pp. 393–431.

Liu, Y., Tsyvinski, A., 2021. Risks and returns of cryptocurrency. The Review of Financial Studies 34 (6), 2689–2727.

Liu, Y., Tsyvinski, A., Wu, X., 2022. Common risk factors in cryptocurrency. The Journal of Finance 77 (2), 1133–1177.

Ljungqvist, A., 2007. IPO underpricing. In: Eckbo, B. E. (Ed.), Handbook of Empirical Corporate Finance. Elsevier, San Diego, Ch. 7, pp. 375–422.

Ljungqvist, A., Nanda, V., Singh, R., 2006. Hot markets, investor sentiment, and IPO pricing. The Journal of Business 79 (4), 1667–1702.

Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. The Journal of Finance 69 (4), 1643–1671.

Loughran, T., McDonald, B., 2016. Textual analysis in accounting and finance: A survey. Journal of Accounting Research 54 (4), 1187–1230.

Loughran, T., Ritter, J. R., 2002. Why don't issuers get upset about leaving money on the table in IPOs? The Review of Financial Studies 15 (22), 413–444.

Luchtenberg, K. F., Seiler, M. J., Sun, H., 2019. Listing agent signals: Does a picture paint a thousand words? The Journal of Real Estate Finance and Economics 59 (617–648), 617–648.

Lyandres, E., Palazzo, B., Rabetti, D., 2019. Do tokens behave like securities? An anatomy of initial coin offerings. Working paper 3287583, Social Science Research Network.

Lyandres, E., Palazzo, B., Rabetti, D., 2022. Initial coin offering (ICO) success and post-ICO performance. Management Science 68 (12), 8515–9218.

Mac an Bhaird, C., Lucey, B., 2010. Determinants of capital structure in Irish SMEs. Small Business Economics 35, 357–375.

Makarov, I., Schoar, A., 2020. Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics 135 (2), 293–319.

Markheim, M., Berentsen, A., 2021. Real estate meets blockchain: Opportunities and challenges of tokenization of illiquid assets. Zeitschrift für Immobilienökonomie 7, 59–80.

Markowitz, H. M., 1952. Portfolio selection. The Journal of Finance 7, 77–91.

Marotta-Wurgler, F., 2008. Competition and the quality of standard form contracts: The case of software license agreements. Journal of Empirical Legal Studies 5 (3), 447–475.

Marotta-Wurgler, F., Chen, D. L., 2012. Does contract disclosure matter? [with comment]. Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft 168 (1), 94–123.

Martin, K. D., Borah, A., Palmatier, R. W., 2017. Data privacy: Effects on customer and firm performance. Journal of Marketing 81 (1), 36–58.

Martin, N., Matt, C., Niebel, C., Blind, K., 2019. How data protection regulation affects startup innovation. Information System Frontiers 21, 1307–1324.

McLaughlin, G. H., 1969. SMOG grading: A new readability formula. Journal of Reading 12 (8), 639–646.

Merello, P., Barberá, A., la Poza, E. D., 2022. Is the sustainability profile of FinTech companies a key driver of their value? Technological Forecasting and Social Change 174, 121290.

Miglo, A., 2021. STO vs. ICO: A theory of token issues under moral hazard and demand uncertainty. Journal of Risk and Financial Management 14 (6).

Miller, A. R., Tucker, C., 2009. Privacy protection and technology diffusion: The case of electronic medical records. Management Science 55 (7), 1077–1093.

Mills, D., Wang, K., Malone, B., Ravi, A., Marquardt, J., Chen, C., Badev, A., Brezinski, T., Fahy, L., Liao, K., Kargenian, V., Ellithorpe, M., Ng, W., Baird, M., 2016. Distributed ledger technology in payments, clearing, and settlement. Finance and Economics Discussion Series 2016-095Washington: Board of Governors of the Federal Reserve System.

Mohan, J., Wasserman, M., Chidambaram, V., 2019. Analyzing GDPR compliance through the lens of privacy policy. In: Gadepally, V., Mattson, T., Stonebraker, M., Wang, F., Luo, G., Laing, Y., Dubovitskaya, A. (Eds.), Heterogeneous Data Management, Polystores, and Analytics for Healthcare. Lecture Notes in Computer Science. DMAH 2019, Poly 2019, Cham, pp. 82–95.

Mohan, V., 2022. Automated market makers and decentralized exchanges: A DeFi primer. Financial Innovation 8 (20), 1–48.

Mollick, E., 2014. The dynamics of crowdfunding: An exploratory study. Journal of Business Venturing 29 (1), 1–16.

Momtaz, P. P., 2020. Initial coin offerings. PLoS ONE 15 (5), e0233018.

Momtaz, P. P., 2021a. The pricing and performance of cryptocurrency. The European Journal of Finance 27 (4–5), 367–380.

Momtaz, P. P., 2021b. Token Offerings Research Database (TORD). https://www.paulmomtaz.com/data/tord, accessed 2 December 2021.

Momtaz, P. P., 2023. Security tokens. In: Baker, H. K., Benedetti, H., Nikbakht, E., Smith, S. S. (Eds.), The Emerald Handbook on Cryptoassets: Investment Opportunities and Challenges. Emerald Publishing Limited, Bingley, pp. 61–78.

Momtaz, P. P., Nam, R. J., Fisch, C., 2022. Blockchain investors. Working Paper 4163004, Social Science Research Network.

Momtaz, P. P., Rennertseder, K., Schröder, H., 2019. Token offerings: A revolution in corporate finance? Working Paper 3346964, Social Science Research Network.

Mood, A. M., 1950. Introduction to the Theory of Statistics. MacGraw-Hill Book Co., New York, pp. 394–398.

Moritz, A., Block, J., Lutz, E., 2015. Investor communication in equity-based crowdfunding: a qualitative-empirical study. Qualitative Research in Financial Markets 7 (3), 309–342.

Mulder, T., Tudorica, M., 2019. Privacy policies, cross-border health data and the GDPR. Information & Communications Technology Law 28 (3), 261–274.

Müller, N. M., Kowatsch, D., Debus, P., Mirdita, D., Böttinger, K., 2019. On GDPR compliance of companies' privacy policies. In: Ekstein, K. (Ed.), Text, speech, and dialogue. TSD 2019. Lecture notes in computer science 11697. Springer, Cham, pp. 151–159.

Nakamoto, S., 2008. Bitcoin: a peer-to-peer electronic cash system. Working paper, https://bitcoin.org/bitcoin.pdf.

Nowak, A., Smith, P., 2016. Textual analysis in real estate. Journal of Applied Econometrics 32 (4), 896–918.

O'Donoghue, T., Rabin, M., 2000. The economics of immediate gratification. Journal of Behavioral Decision Making 13 (2), 233–250.

OECD, 2020. The tokenisation of assets and potential implications for financial markets. OECD Blockchain Policy Series.

Ofir, M., Sadeh, I., 2020. ICO vs. IPO: Empirical findings, information asymmetry, and the appropriate regulatory framework. Vanderbilt Journal of Transnational Law 53 (2), 525–614.

Omri, B.-S., Schneider, C. E., 2014. More than you wanted to know: The failure of mandated disclosure. Princeton University Press, Princeton.

Pai, A., Geltner, D., 2007. Stocks are from Mars, real estate is from Venus. The Journal of Portfolio Management 33 (5), 134–144.

Papke, L. E., Wooldridge, J. M., 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. Journal of Applied Econometrics 11 (6), 619–632.

Parlour, C. A., Rajan, U., Zhu, H., 2022. When FinTech competes for payment flows. The Review of Financial Studies 35 (11), 4985–5024.

Pazos, J., 2019. Valuation method of equity-based security token offerings (STO) for start-up companies. The Journal of The British Blockchain Association 2 (1), 1–10.

Peacock, C., Milewicz, K., Snidal, D., 2019. Boilerplate in international trade agreements. International Studies Quarterly 63 (4), 923–937.

Peterson, K., Schmardebeck, R., Wilks, T. J., 2015. The earnings quality and information processing effects of accounting consistency. The Accounting Review 90 (6), 2483–2514.

Porter, M. E., 1998. Clusters and the new economics of competition. Harvard Business Review 76 (6), 77–90.

Posner, R. A., 1981. The economics of privacy. The American Economic Review 71 (2), 405–409.

Puschmann, T., 2017. Fintech. Business & Information Systems Engineering 59 (1), 69–76.

Ramadorai, T., Uettwiller, A., Walther, A., 2021. The market for data privacy. Working Paper 3352175, Social Science Research Network.

Ritter, J. R., 1984. The "hot issue" market of 1980. The Journal of Business 57 (2), 215–240.

Ro, S., Ziobrowski, A. J., 2012. Wealth effects of REIT property-type focus changes: Evidence from property transactions and joint ventures. Journal of Property Research 29 (3), 177–199.

Rock, K., 1986. Why new issues are underpriced. Journal of Financial Economics 15 (1), 187–212.

Roosenboom, P., van der Kolk, T., de Jong, A., 2020. What determines success in initial coin offerings? Venture Capital 22 (2), 161–183.

Saleh, F., 2021. Blockchain without waste: Proof-of-stake. The Review of Financial Studies 34 (3), 1156–1190.

Salton, G., Wong, A., Yang, C. S., Nov. 1975. A vector space model for automatic indexing. Commun. ACM 18 (11), 613–620.

Schueffel, P., 2016. Taming the beast: A scientific definition of fintech. Journal of Innovation Management 4 (4), 32–54.

Schultz, P. H., Zaman, M. A., 1994. Aftermarket support and underpricing of initial public offerings. Journal of Financial Economics 35 (2), 199–219.

Schweizer, D., Zhou, T., 2017. Do principles pay in real estate crowdfunding? The Journal of Portfolio Management 43 (6), 120–137.

Schär, F., 2021. Decentralized finance: On blockchain- and smart contract-based financial markets. Federal Reserve Bank of St. Louis Review 103 (2), 153–174.

Schär, F., Berentsen, A., 2020. Bitcoin, Blockchain, and Cryptoassets: A Comprehensive Introduction. MIT Press, Cambridge.

Seiler, M., Madhavan, P., Liechty, M., 2012. Toward an understanding of real estate homebuyer internet search behavior: An application of ocular tracking technology. Journal of Real Estate Research 34, 212–242.

Shahrokhi, M., Parhizgari, A. M., 2020. Crowdfunding in real estate: evolutionary and disruptive. Managerial Finance 46 (6), 785–801.

Signori, A., Vismara, S., 2018. Does success bring success? The post-offering lives of equity-crowdfunded firms. Journal of Corporate Finance 50, 575–591.

Singh, R. I., Sumeeth, M., Miller, J., 2011. A user-centric evaluation of the readability of privacy policies in popular web sites. Information Systems Frontiers 13, 501–514.

Sirmans, G. S., Macpherson, D. A., Zietz, E. N., 2005. The composition of hedonic pricing models. Journal of Real Estate Literature 13 (1), 3–43.

Smart, S. B., Thirumalai, R. S., Zutter, C. J., 2008. What's in a vote? The short- and long-run impact of dual-class equity on IPO firm values. Journal of Accounting and Economics 45 (1), 94–115.

Sockin, M., Xiong, W., 2023. Decentralization through tokenization. The Journal of Finance 78 (1), 247–299.

Spence, M., June 2002. Signaling in retrospect and the informational structure of markets. American Economic Review 92 (3), 434–459.

Stewart, H., Jürjens, J., 2018. Data security and consumer trust in fintech innovation in Germany. Information and Computer Security 26 (1), 109–128.

Stoughton, N. M., Zechner, J., 1998. IPO-mechanisms, monitoring and ownership structure. Journal of Financial Economics 49 (1), 45–77.

Strahilevitz, L. J., Kugler, M. B., 2016. Is privacy policy language irrelevant to consumers? The Journal of Legal Studies 45 (S2), 69–95.

Sunyaev, A., Dehling, T., Taylor, P. L., Mandl, K. D., 2015. Availability and quality of mobile health app privacy policies. Journal of the American Medical Informatics Association 22 (1), 28–33.

Swinkels, L., 2023. Empirical evidence on the ownership and liquidity of real estate tokens. Financial Innovation 9 (1), 1–29.

Tao, R., Su, C.-W., Naqvi, B., Rizvi, S. K. A., 2022. Can fintech development pave the way for a transition towards low-carbon economy: A global perspective. Technological Forecasting and Social Change 174, 121278.

Tapscott, A., Tapscott, D., 2017. How blockchain is changing finance. Harvard Business Review 1 (9), 2–5.

Tesfay, W. B., Hofmann, P., Nakamura, T., Kiyomoto, S., Serna, J., 2018. PrivacyGuide: Towards an implementation of the EU GDPR on internet privacy policy evaluation. In: IWSPA '18 (Ed.), Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics. ACM, New York, pp. 15–21.

Thaler, R. H., Johnson, E. J., 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. Management Science 36 (6), 643–660.

Thewissen, J., Thewissen, J., Torsin, W., Arslan-Ayaydin, Ö., 2022. Linguistic errors and investment decisions: the case of ICO white papers. The European Journal of Finance forthcoming, 1–43.

Tsai, J. Y., Egelman, S., Cranor, L., Acquisti, A., 2011. The effect of online privacy information on purchasing behavior: An experimental study. Information Systems Research 22 (2), 254–268.

Uzsoki, D., 2019. Tokenization of Infrastructure: A blockchain-based solution to financing sustainable infrastructure. International Institute for Sustainable Development.

Vismara, S., 2019. Sustainability in equity crowdfunding. Technological Forecasting and Social Change 141, 98–106.

Vulkan, N., Åstebro, T., Sierra, M. F., 2016. Equity crowdfunding: A new phenomena. Journal of Business Venturing Insights 5, 37–49.

Wachter, S., 2018. The GDPR and the Internet of Things: A three-step transparency model. Law, Innovation and Technology 10 (2), 266–294.

Weesie, J., 1999. Seemlingly unrelated estimation and the cluster-adjusted sandwich estimator. Stata Technical Bulletin 9 (52), 231–248.

Welch, B. L., 1947. The generalization of "student's" problem when several different population variances are involved. Biometrika 34 (1-2), 28–35.

Welch, I., 1989. Seasoned offerings, imitation costs, and the underpricing of initial public offerings. The Journal of Finance 44 (2), 421–449.

Wild, F., 2007. An LSA package for R. In: Wild, F., Kalz, M., van Bruggen, J., Koper, R. (Eds.), Mini-Proceedings of the 1st European Workshop on Latent Semantic Analysis in Technology-Enhanced Learning. Heerlen, pp. 11–12.

Wolff, J., Atallah, N., 2021. Early GDPR penalties: Analysis of implementation and fines through may 2020. Journal of Information Policy 11 (3748837), 63–103.

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.

World Bank, 2021. World Development Report 2021: Data for Better Lives. World Bank, Washington, D.C.

Wright, A., De Filippi, P., 2015. Decentralized blockchain technology and the rise of lex cryptographia. Working Paper 2580664, Social Science Research Network.

Yermack, D., 2015. Is bitcoin a real currency? An economic appraisal. In: Chuen, D. L. K. (Ed.), Handbook of Digital Currency. Academic Press, San Diego, pp. 31–43.

Yermack, D., 2017. Corporate governance and blockchains. Review of Finance 21 (1), 7–31.

Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association 57 (298), 348–368.

Zhang, Y., Wang, T., Hsu, C., 2020. The effects of voluntary GDPR adoption and the readability of privacy statements on customers' information disclosure intention and trust. Journal of Intellectual Capital 21 (2), 145–163.

Zheng, S. X., Li, M., 2008. Underpricing, ownership dispersion, and aftermarket liquidity of IPO stocks. Journal of Empirical Finance 15 (3), 436–454.

Zhu, C., 2019. Big Data as a governance mechanism. The Review of Financial Studies 32 (5), 2021–2061.