## Table of Contents

# 1. Introduction

“The next generation for markets, the next generation for securities, will be tokenization of securities and real-world assets,” declared Larry Fink, the CEO of BlackRock. This is endorsed by David Solomon, the CEO of Goldman Sachs, who stated that the use of decentralized finance would help mitigate risks and increase transparency in the financial system. Therefore, a trustworthy tokenization process of financial instruments is rightfully expected to revolutionize the financial landscape.

Tokenization is a transformative process that converts assets, services, or activities into singular identifiers that represent digital certificates of ownership. The tokenized instruments are linked to digital tokens, which are created and built on a distributed ledger technology (DLT) infrastructure or blockchain technology, allowing individuals to buy, sell and trade the tokens as units. Each token has its own programming code and set of rules which ensure immutability, identifiability, and traceability. Everything, ranging from real estate to art, and even stars, can be tokenized. However, it is crucial to consider that this comes with its risks and challenges, which will be thoroughly analyzed further on in this section.

To start the tokenization process, the first step is to build a strong ecosystem for each specific token. This means dealing with both the technical and legal aspects to be able to satisfy the parties’ needs. Next, the configuration of the token is crucial. Regardless of the technology used, the coding and programming behind the token’s representation should be clear and intuitive. Moreover, for the investor to be persuaded that the token value represents a good investment opportunity, the issuer will need to provide extensive information about the expected revenues and the rights associated to the token in question. Note that each token owns a value which can be computed using the following formula:

$$\text{Token value} = \frac{\text{total value of the asset} \times \text{percentage of tokenized assets}}{\text{number of issued tokens}}$$

Once the token is set up, the final step is to transfer it to the right recipients, and this ending procedure is called token distribution. At this point, the course of action is the following: the involved investor sends the funds to a body in charge, which asks the custodian of the tokens (a trusted central authority) to transfer them to the investor's wallet. Once the investor secures the token, the body responsible for the distribution process releases the payment of the funds to the issuer. This is a faster and cheaper process compared to the traditional procedure of distribution of assets and securities. In fact, specifically in the financial sector, tokenization is expected to have a substantial impact on markets. This is because, as previously mentioned, everything can be tokenized: traditional assets and securities, both physical and digital, ranging from cash to equities to real estate. In the context of this paper, note that a tokenized security indicates the tokens of a stock, bond, or investment fund.

There are several advantages related to tokenization, with one of the most significant being to divide the assets and securities that are non-fungible in the traditional world. Through tokenization, their digital ownership can be shared across multiple people, leading to a better-connected financial market. Thus, the so-called fractional ownership lowers the barrier to investments, fostering financial inclusion and contributing to the democratization of investment opportunities. It promotes a global access to these opportunities, making them accessible even to investors with smaller portfolio sizes and lower minimum tickets. Fractional ownership also allows for a quicker trading of the tokens, removing the time to wait for profits and losses of large illiquid assets, thus leading to an increased liquidity compared to markets that trade assets and securities only in large clusters. Moreover, the use of blockchain applications and DLT infrastructures eliminates most middlemen in secondary markets, fostering liquidity even further. Even though not all middlemen are removed (some, such as distributors, are still required), the tokenization process reduces the transaction and process time while also lowering middlemen fees. According to the statistical analyses of Cashlink in 2020, when comparing the established private placement and the innovative blockchain technology, the latter reduces bond issuance costs by up to 90 percent and costs of fundraising by almost 40 percent (Benedetti and Rodriguez-Guarnica, 2022). Plus, this becomes even more of an advantage when looking at empirical data: tokenization offers an alternative means of financial services to those who don’t have access to bank accounts, which represent 30% of adult people according to the World Bank’s Global Findex Database of 2021. By bypassing intermediary banks, tokenization also accounts itself as more transparent and efficient. In fact, it being an automated process would significantly reduce the number of human errors and oversights, thus saving time and costs and providing more accuracy. The back-office administrative work would also be significantly reduced, thus lowering clearing and settlement times as well.

It is through the union of tokenization and blockchain that most advantages arise. In fact, anyone using the blockchain interface for a particular financial instrument can access its history of transactions, ensuring a stable market with fewer manipulations, making investors feel safer. The fact that all transactions are recorded on a public ledger reduces the potential risk of information disputes. Data security is also enforced by the blockchain’s intrinsic cryptographic functions. As a matter of fact, according to the General Data Protection Regulation (GDPR), the blockchain ensures the unicity of each token: any individual who possesses the alphanumeric code of the token would only be able to refer to the token, but not to access any of the sensitive data stored inside.

It is important to note that tokenization can also benefit broader goals, such as sustainability. In fact, the tokenization of green bonds, renewable energy, or environmental projects offers the investors the opportunity to support ventures that prioritize a more sustainable development. Thus, global causes such as sustainability can be seen as an additional incentive for adopting this new technology for investment purposes.

To resume, the spreading of the tokenization process would lead to several key advantages: transparency, efficiency, immutability, security, lower costs, enhanced liquidity, and even more sustainable investments.

On the other hand, as previously mentioned, despite its numerous benefits, the tokenization process also presents a few challenges that need to be addressed. One of the primary issues regarding tokenization is security, as the system would need to interoperate solidly to ensure the integrity of the financial instruments. Even though blockchain demonstrated to be effective in terms of safety, if tokenization is adopted on a global scale, a wide number of technical challenges related to the scalability of the underlying technology need to be faced. For instance, the risks of cybersecurity attacks must be addressed and minimized. Investors need to feel that the infrastructure is stable and reliable enough to adopt it over traditional methods. The system is thus required to have appropriate levels of privacy and security, thus minimizing the risks of cybersecurity hacks, market manipulations and frauds. In addition to that, it must be considered that general hesitance and an overall lack of knowledge in front of this new technology could hold investors back from adopting it.

More importantly, it goes without saying that a process such as tokenization is extremely complex and requires a remarkable number of regulations. Across the various jurisdictions, the regulatory framework regarding tokenization is still under discussion and development. This could represent a problem for investors, who might be discouraged by the overall uncertainty. As of now, there is an established financial testing ground in the United Kingdom and in Spain. In Europe, the DLT-Pilot regulation (the European Union's Regulation 2022/858) became effective on March 23rd, 2023. It introduced a European regulatory sandbox, providing harmonized requirements to all the market participants wanting to be authorized to get a DLT market infrastructure. With privacy regimes such as the GDPR and especially its Data Erasure clause, it becomes harder to address the regulatory issue posed by tokenization. In fact, clauses like Data Erasure allow individuals the so-called “right to be forgotten”, which completely contradicts the immutability and traceability attributes of the blockchain. Even though there are efforts that encourage the establishment of tokenization, it remains that there is no current single universal established regulatory landscape. Plus, there are still divergences on how different jurisdictions are dealing with this new technology: some proactively advocate the tokenization of financial instruments while others remain skeptical. Since stakeholders request regulatory clarity before they even consider investing in a new technology, regulators must get to work to homogenize the present legal framework.

Fraud, manipulation, hacking, and the lack of a universally accepted regulatory landscape are challenges to be tackled but that do not make the tokenization process unfeasible. In fact, collaboration with the appropriate stakeholders, especially the regulators, would be the key to implement a standardized set of directives regarding security and customer protection. This harmonization of the implementation of the tokenization is definitely necessary and highlights the measureless work that needs to be carried out before mass adoption occurs.

Just recently, a panel at the University of Pennsylvania 2024 Blockchain Conference shed light on how Traditional Finance is embracing crypto, providing some extremely valuable and up-to-date insights on tokenization. Oliver Harris (Head of

Digital Assets at Goldman Sachs), Sandy Kaul (Senior Vice President at Franklin Templeton), Pranav Kanade (Head of Digital Assets at Van Eck) and Joshua Ashley Klayman (U.S. Head of FinTech and Head of Blockchain & Digital Assets at Linklaters) shared their views on the evolution of the process of tokenization, along with the most common misconceptions.

First, the panel highlighted that there is a foundational lack of understanding of the crypto space. The main issues are that people fail to understand that commerce can take place in protocols without intermediaries and many still believe that bitcoin is untraceable and that those involved in the blockchain space are completely uninterested in compliance. Surprisingly, the panel emphasized that it would be easy to address most of these misconceptions if they were compact, but it becomes harder and harder when they are spread around. Since tokenization represents a completely new protocol economy, there should be more education about the type of blockchain and the types of assets and securities an entity is issuing.

Moreover, it was strongly underlined that the infrastructure on which securities rely on was built in the early 1970s and has not been updated to any significant degree since then. This means that the system has fundamentally been operating in the same way for over fifty years. Nonetheless, this represents a tremendous opportunity to replace the entire financial market infrastructure at a global level, thus fully changing the way financing works all over the world.

The panel also discussed the credibility that Exchange-Traded Funds (ETFs) are bringing into the traditional financing world. In fact, the advent of ETFs is seen as the legitimization of the idea that there is a completely new pool of opportunities: ETFs constitute the evidence of the existence of a totally new asset class.

Finally, they underscored the benefits tokenization would bring, from the revenue opportunities to the minimization of the operational risk. As previously described in this section, being able to issue a token on the blockchain would be vastly democratizing. In fact, there's this increasingly growing thought that even smaller corporations would issue bonds on their own. In addition to that, capital raising would be executed directly by going to the community, thus forging closer relationships. This represents the disruption of the conventional idea that the only way to own a company or a part of it is through equity. The panel concluded by noting that this would also change the way big corporations think, as tokenization brings up the possibility of tokenizing not just assets and securities, but even much more complex resources such as algorithms. There is a truly limitless potential in the field.

The tokenization of financial instruments would bring a wide range of innovative investment opportunities, facilitating and enhancing financial inclusion. This paper aims to discuss the impact that the very possible future tokenization of securities will have on financial markets, and it includes an empirical analysis that forecasts this effect. Its objective is to test a measure of liquidity on current non-tokenized securities' data and subsequently estimate a benchmark for tokenized securities. The paper is structured as follows. To start, section 2 extensively tackles the related literature. In particular, given the study on Stablecoins, it will be reasonable to suppose that one of the major benefits of the implementation of the tokenization process on securities is an enlarged liquidity. Consequently, the main hypothesis of the empirical analysis covered in Section 3 is formulated as follows:

Research hypothesis. *The tokenization of securities significantly influences the liquidity metrics of the financial market*.

The presence of a higher liquidity would benefit investors in multiple ways, from having a more efficient market and thus reducing transaction costs, to improving price stability and thus promoting investments with fewer price movements. Section 3 begins with an explanation of how the data collection was carried out, the actual method and model design, and ends with a discussion of its results. Section 4 concludes.

## 2. Related Literature

Since the tokenization process is still in its early development, the related literature, especially the one strictly regarding securities, is only currently emerging, even though it is seen as the one with the most immediate potential growth. Previous research mainly analyzes the functioning of tokenization and the implications it had on the assets it has been implemented on. In fact, existing analyses focus either on methods to describe a token or on generic token applications. Most of the papers regarding the first stream neglect the investigation of the implications of the process of tokenization. Only a few empirical analyses have been carried out on the effects, benefits, and drawbacks of the implementation of digital currencies. This section concentrates on the application of tokens since the established literature is still vague about how they can be applied in a specific framework. This includes addressing the benefits mentioned earlier, such as reduced investment costs, increased liquidity, and enhanced transparency. To gather more information for the empirical analysis carried out further on in this paper, studies on tokenized cash, specifically Stablecoins, have been considered.

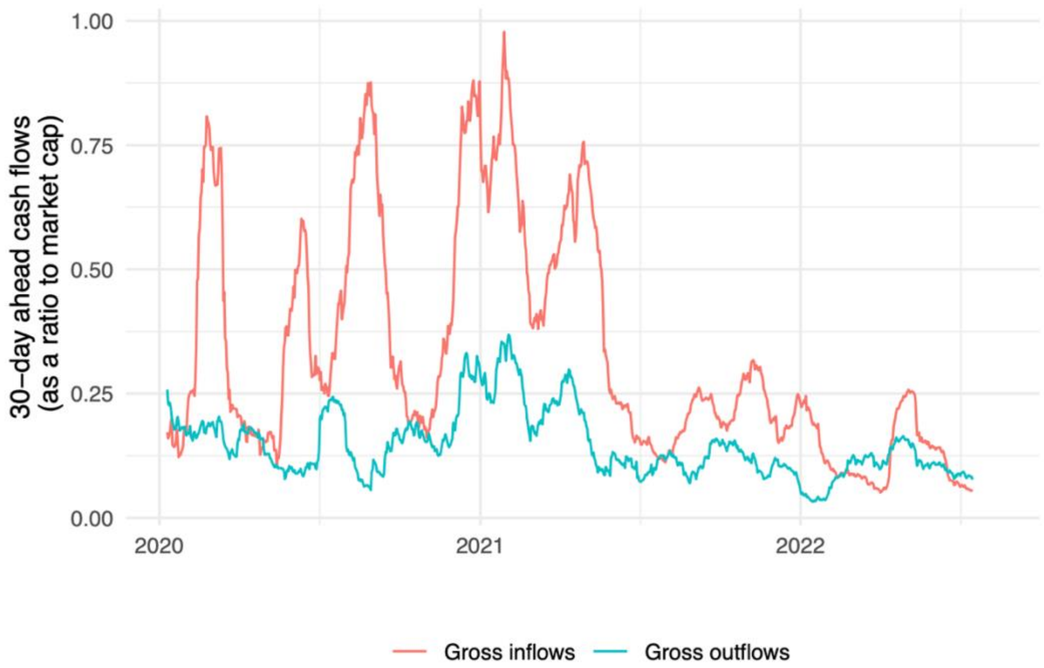
Stablecoins can range from tokenized cash, which are cryptographically secured through a DLT infrastructure and are backed by an underlying reserve of cash, to algorithmic Stablecoins, which are the most controversial since they rely on themselves without being backed by any external asset.

Stablecoins are tokenized assets that peg their value to some external monetary reference, which usually is the U.S. Dollar (USD) but can also be other fiat currencies, cryptocurrencies, or commodities like gold. Most of the Stablecoins come with the promise that they can be bought back at par upon request, such that investors consider their value stable when compared to fiat currency. In the cryptocurrency ecosystem, they are an essential foundation for decentralized finance protocols, thus enabling lending and asset exchange. There are five major Stablecoins that circulate since the beginning of 2022: USD Coin (USDC) and Binance USD (BUSD), which hold tokenized cash in their reserves; Tether (USDT), which is fiat asset-backed and which we will come back to further on in this paper; Terra (UST), which is an algorithmic Stablecoin that collapsed in 2022; and Dai (DAI), a crypto-backed coin. Stablecoins have been chosen as an example in this paper because it is believed they can widely substitute cash in traditional banking and more generally the standard means of payment, in businesses and even in households. In fact, since their first adoption in 2020 on public blockchains like Ethereum (the most popular blockchain that comes with the largest ecosystem), Stablecoins showed an enormous growth. They seem to offer a great potential, with around $130 billion increasing supply as of October 2021 (U.S. Department of the Treasury, 2021). Although it remains a relatively small amount compared to the vast scale of more than $4 trillion money market fund complex (Liao, 2022), it still shows a close to 500% increase over the preceding year (Liao and Caramichael, 2022).

Stablecoins are currently used because of the benefits mentioned in Part 1: investors prefer digital currencies because of the nearly instantaneous 24/7 trading without intermediaries and with lower fees. Moreover, as noted by Liao in "Macroprudential Considerations for Tokenized Cash" in 2022, a broad and diversified ownership of tokenized assets is shown by the fact that approximately 75% of the wallets that hold tokenized cash have balances below $100 (Liao, 2022).

Now, focusing on the USD Coin, it is shown below that gross inflows overshadowed outflows. Both cash flows are computed forward-looking. The graph represents the 30-day ahead USDC gross inflows and outflows as a ratio of market capitalization outstanding. In most of the analyzed period, gross inflows are remarkably higher than gross outflows, which marks the increasing demand for tokenized cash. Note that the maximum difference between the two flows reached 36.9%, where gross inflows were about three times gross outflows.

*Figure 1: Outflows and Inflows (Liao, 2022)*




Another point that was highlighted in Section 1 was that tokenized financial instruments should lead to greater liquidity compared to regular assets and securities. To prevent liquidity-induced runs, banking regulatory frameworks adopted the Basel III Liquidity Coverage Ratio (LCR), which evaluates the quality of the assets over the potential risks of cash outflows over the next 30 calendar days. The ratio is required to be over 100% for banks, and below is the formula:

$$\frac{\text{Stock of HQLA}}{\text{Total net cash outflows over the next 30 calendar days}} \geq 100\%$$

where

$$\text{Net cash outflows} = \text{cash outflows - min}\left(\text{cash inflows, } 0.75 \times \text{cash outflows}\right)$$

The stock of HQLA in the numerator stands for the quality and liquidity of the assets, while the denominator reflects the perceived runoff risk associated with liability items. The stock of HQLA represents all the liquid assets a bank possesses that can quickly be transformed into cash in cases of stress. There are two levels of HQLAs, based on liquidity and devaluation: Level 1 HQLAs, which are the most liquid and aren't affected by any devaluation, and Level 2 HQLAs, which conversely are less liquid and undergo 15% devaluation. Gordon Y. Liao (2022) proposes an empirical analysis in his study in which he applies the Basel III LCR to cash tokens, employing three different approaches depending on the runoff rates. First, considering all token holders as non-operational deposits, thus applying a runoff rate of 40%, a liquidity ratio of 196% was found. Then, assuming the worst observed runoff rate, the LCR sums up to 850%. At last, employing a conservative approach which assumes the worst gross outflows and no expectation of inflows, the liquidity ratio amounts to 212%. Through this study, it is found that liquidity ratios range from 196% to 850%, all of which exceed the 118% liquidity ratio computed for the eight Systemically Important Banks in the U.S. in Q2 2022. Therefore, despite the serious consumer protection, market safety challenges, and the fact that Stablecoins are still in their infancy, they appear to show an immense potential for innovation.

Closely looking at securities instead, the future for tokenization looks prosperous as well, but it is harder to find concrete examples on how the tokenization has been implemented. Having said that, as an example, in the UK Financial Conduct Authority (FCA) testing ground briefly mentioned in Section 1, the first company to issue a tokenized bond was Nivaura, a fintech company. Nivaura's project dealt with the issuance of a Control bond, a GDP-denominated bond which relates to the tokenization of fiat currency, and an Experimental bond, an Ethereum-denominated bond which uses smart contracts and constitutes the first cryptocurrency bond issued on an open public blockchain (Cohen et al., 2018). The issuance of the Control bond demonstrated enhanced transparency and accuracy in owning the assets, since the transfer of the token directly happened between the issuer and the investor, thus eliminating the need for intermediaries, and allowing the transaction to be publicly and permanently recorded on the blockchain. The issuance of the Experimental bond, through smart contracts and peer-to-peer trading, represents the evidence of how much the costs and complexity of the process are reduced. Even according to Santander Bank, when they issued a bond end-to-end on a public blockchain in September 2019, the benefits were met, showing enhanced security and market transparency, reduced costs, and reduced settlement times (Benedetti and Rodriguez-Garnica, 2022).

In today's world, securities are abundant, but it is challenging to physically divide and transfer them, and investors currently choose to trade unsecured files that represent the whole financial instrument and sometimes even only just a part of it. Thanks to the tokenization process, the securities' rights would be almost instantaneously transferrable through peer-to-peer trading, thus managing to avoid these cumbersome systems. Moreover, following the example of Stablecoins, the tokenization of securities is also expected to reach higher values of liquidity.

# 3. Empirical Analysis

## a) Data Collection

Since tokenization is a relatively recent process especially for securities, no comprehensive database containing information about tokenized securities exists yet. Therefore, the dataset of 7 columns and 6122 observations used in this study is composed of information about non-tokenized securities. It comprises an overall view of Nvidia Corporation (NVDA)'s stocks. Nvidia is a leading technology company, known for its cutting-edge high-end GPUs, innovative AI hardware and software. The stock of the company is traded on major exchanges with the ticker symbol NVDA, and it is sought-after as one of the most prominent investments in the technology field. The dataset was found on Kaggle and it contains Nvidia's historical stock prices and performance of all trading days from January 1999 to May 2023. It includes the following intuitively named variables:

- "Date," the trading day,
- "Open," the opening price of Nvidia's stock at the starting period of a trading session on a security exchange,
- "High," the highest price at which a stock is traded during a specific trading day,
- "Low," the lowest price at which a stock is traded in a trading session, which is typically lower than the opening or closing price,
- "Close," the last price at which a stock is traded during a trading session,
- "Adj Close," the adjusting closing price of a stock that accounts for possible corporate actions, such as stock splits or dividends,
- "Volume," the total number of stocks traded during a trading session, which usually is an indicator of market strength. The terms "Volume" and "Trading Volume" will be used interchangeably in this paper.

This dataset, focusing on non-tokenized securities, doesn't contain a feature that could serve as a proxy for tokenization. Due to this limitation, an alternative dataset, about the price of the Stablecoin cryptocurrency Tether (one of the five major Stablecoins, mentioned earlier) was identified. The Tether tokens of this dataset are issued from Tether Limited, a company based in Hong Kong, and the chosen dataset for this study includes data for all trading days from November 9$^{th}$, 2017, to May 24$^{th}$, 2022. The Tether dataset includes the same 7 columns ("Date," "Open," "High," "Low," "Close," "Adj Close," and "Volume") as the Nvidia dataset. It has fewer observations, totaling 1658. In the Tether dataset, the feature we are the most interested in studying is "Volume," which here too represents the amount of stock traded in the trading session. We will consider this variable as a proxy for tokenization in our empirical analysis. To accomplish this, we merged the Nvidia dataset with the "Volume" column of the Tether dataset, which was renamed "Tether_Volume" to avoid confusion with the existing "Volume" column coming from the Nvidia dataset. However, since the Tether dataset contained a fewer number of observations corresponding to a smaller number of days, the merge process was restricted to those specific days.

**b) Research Methodology**

Through the analysis of previous research on tokenized cash, specifically Stablecoins, it was found that increased liquidity is one of the major benefits of the implementation of the tokenization process. Therefore, as previously mentioned in the hypothesis, the empirical analysis that follows in this paper is expected to bring the same results. While in the Stablecoins study the Basel III Liquidity Coverage Ratio (LCR) was employed to give a measure of liquidity, it goes without saying that we cannot use the value of the stock of High-Quality Liquid Assets (HQLA) to evaluate the measure of liquidity for securities. Fortunately, there are several other ways to explore liquidity metrics, and given our variables, the price impact ratio was first calculated in this case. The price impact ratio indicates how much the price changes in a market in response to an incoming order (Bouchaud, 2009), namely when a trader buys or sells a financial instrument. For the purpose of this research, the incoming order was defined as the percentage change in the volume of the assets. Thus, the price impact variable was computed through the following formula:

$$\text{Price Impact} = \frac{\%\text{ change in Price}}{\%\text{ change in Volume}}$$

For each trading day, the percentage change in price was calculated using the closing price, while the percentage change in volume was determined through the volume, as follows:

$$\%\text{ change in Price} = \frac{\text{Closing Price Day}_0 - \text{Closing Price Day}_{-1}}{\text{Closing Price Day}_{-1}} \times 100$$

$$\%\text{ change in Volume} = \frac{\text{Volume Day}_0 - \text{Volume Day}_{-1}}{\text{Volume Day}_{-1}} \times 100$$

A higher value of the price impact would show that changes in the stock volume have a larger impact on the price of securities, thus indicating lower liquidity. The price impact for the non-tokenized securities of the Nvidia dataset averaged -0.2045. The result is very close to zero, thus indicating that the volume of the stock traded doesn’t seem to have a big effect on the prices of the stock. This indicates moderate to high liquidity. However, this analysis is not enough to assess liquidity, thus it is necessary to relate the price impact ratio to other liquidity indicators, such as the bid-ask spread. Since the bid-ask spread variable isn’t contained in the dataset, we proceed by performing feature engineering. As its name says, the bid-ask spread refers to the difference between the ask price and the bid price in financial markets. For situations where quote data is unavailable, such as in our case, previous research (Abdi & Ranaldo, 2017) proposed a new method to evaluate the bid-ask spread through readily available data that we also find in our dataset: daily high and low prices. Following their easy-to-compute method, we calculated the bid-ask spread through the following formula:

$$\text{Bid-Ask Spread Percentage} = \frac{\text{Daily High - Daily Low}}{2 \times \text{Mid-Range}} \times 100$$

where the mid-range represents the average of daily high and low prices. Moreover, note that the method suggested by Abdi and Ranaldo is presumed to improve accuracy and does not depend on bid-ask bounces. Now that we have the metrics for liquidity, we can move towards the explanation of the training of the empirical models.

The traditional approach to estimate the effect that tokenized securities will have on liquidity would require the gathering of data on price impact and bid-ask spread for tokenized securities. However, as previously mentioned, no universal dataset that compiles data on tokenized securities exists so far. Therefore, another line of action was adopted in this paper. Through machine learning algorithms, we will forecast how the tokenization process may impact the liquidity of securities based on the relationship between liquidity in non-tokenized securities and a variable proxy for tokenization. A final interpretation of the results is required, where a positive and statistically significant coefficient of the proxy for tokenization would show improved liquidity. It will be essential to consistently bear in mind, throughout the duration of the empirical analysis, that there are other factors that influence liquidity besides tokenization, such as investor behavior, the regulatory environment, and market conditions.

In this study, we first implemented a Linear Regression model to evaluate the coefficient and p-value of the variable acting as a proxy for tokenization. Subsequently, we employed two additional machine learning models, namely the Gradient Boosting regression and the Random Forest regression, to further investigate the underlying relationships among the dependent variable and the independent variables. This comprehensive approach allowed us to thoroughly analyze the dataset and gather meaningful results.

### c) Results

The outcomes of the study revealed by the models are further discussed in this part of the paper. First, an exploratory data analysis was performed to check for outliers and missing data on both datasets. As we can see from the data displayed in Figure 2a and Figure 2b, the data demonstrates low variance, suggesting that the variables are all around the same values and indicating the absence of outliers. This is endorsed by the narrow value of the interquartile range (IQR), which is around 25 for each value in the Nvidia dataset, and close to 0 for the Tether dataset. The difference in the values of the two interquartile ranges can be attributed to the wider range of values observed in the Nvidia dataset when compared to the Tether dataset.

*Figure 2a: Description of the data of the Nvidia dataset*

| | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| **count** | 6122.000000 | 6122.000000 | 6122.000000 | 6122.000000 | 6122.000000 | 6.122000e+03 |
| **mean** | 32.083172 | 32.698178 | 31.452291 | 32.105539 | 31.846044 | 6.128430e+07 |
| **std** | 61.979191 | 63.229129 | 60.681098 | 62.022923 | 62.028533 | 4.400809e+07 |
| **min** | 0.348958 | 0.355469 | 0.333333 | 0.341146 | 0.313034 | 1.968000e+06 |
| **25%** | 2.677500 | 2.758750 | 2.603333 | 2.677500 | 2.456865 | 3.438680e+07 |
| **50%** | 4.320000 | 4.412500 | 4.245000 | 4.335000 | 3.981222 | 5.138220e+07 |
| **75%** | 27.708750 | 27.966875 | 27.088750 | 27.643750 | 27.279851 | 7.457340e+07 |
| **max** | 335.170013 | 346.470001 | 320.359985 | 333.760010 | 333.350800 | 9.230856e+08 |

*Figure 2b: Description of the data of the Tether dataset*

| | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| **count** | 1658.000000 | 1658.000000 | 1658.000000 | 1658.000000 | 1658.000000 | 1.658000e+03 |
| **mean** | 1.001815 | 1.008433 | 0.996067 | 1.001816 | 1.001816 | 4.039061e+10 |
| **std** | 0.006087 | 0.010536 | 0.008294 | 0.006034 | 0.006034 | 3.979689e+10 |
| **min** | 0.972522 | 0.978690 | 0.899490 | 0.966644 | 0.966644 | 3.581880e+08 |
| **25%** | 1.000041 | 1.001122 | 0.993818 | 1.000043 | 1.000043 | 4.279052e+09 |
| **50%** | 1.000714 | 1.006298 | 0.998211 | 1.000668 | 1.000668 | 3.027514e+10 |
| **75%** | 1.003138 | 1.012296 | 1.000029 | 1.002973 | 1.002973 | 6.100956e+10 |
| **max** | 1.080950 | 1.105910 | 1.021830 | 1.077880 | 1.077880 | 2.790675e+11 |

Looking at Figure 3, we notice that both datasets present no missing data, which is advantageous and timesaving for our analysis since it enhances the accuracy and reliability of the results and drives to zero the need for imputation techniques.

*Figure 3: Missing data in the Nvidia dataset (left) and in the Tether dataset (right)*

| | Total | Percent |
|---|---|---|
| **Date** | 0 | 0.0 |
| **Open** | 0 | 0.0 |
| **High** | 0 | 0.0 |
| **Low** | 0 | 0.0 |
| **Close** | 0 | 0.0 |
| **Adj Close** | 0 | 0.0 |
| **Volume** | 0 | 0.0 |

| | Total | Percent |
|---|---|---|
| **Date** | 0 | 0.0 |
| **Open** | 0 | 0.0 |
| **High** | 0 | 0.0 |
| **Low** | 0 | 0.0 |
| **Close** | 0 | 0.0 |
| **Adj Close** | 0 | 0.0 |
| **Volume** | 0 | 0.0 |

Then, as previously outlined in the Research Methodology section, we merged the Nvidia dataset with the "Volume" column of the Tether dataset based on the "Date" column of the latter. This operation enabled us to have a proxy for tokenization that we required to carry on the analysis. We can observe the head of the merged dataset in Figure 4.

*Figure 4: Head of the merged dataset*

| | Date | Open | High | Low | Close | Adj Close | Volume | Tether_Volume |
|---|---|---|---|---|---|---|---|---|
| **0** | 2017-11-09 | 51.317501 | 51.582500 | 50.092499 | 51.330002 | 50.768211 | 97856400.0 | 358188000 |
| **1** | 2017-11-10 | 53.270000 | 54.667500 | 52.907501 | 54.035000 | 53.443600 | 125325600.0 | 756446016 |
| **2** | 2017-11-13 | 54.035000 | 54.292500 | 53.002499 | 53.157501 | 52.575703 | 58237600.0 | 746227968 |
| **3** | 2017-11-14 | 53.250000 | 53.700001 | 52.807499 | 53.544998 | 52.958961 | 52929200.0 | 1466060032 |
| **4** | 2017-11-15 | 52.987499 | 53.000000 | 51.950001 | 52.494999 | 51.920467 | 50194800.0 | 767884032 |

Since we constructed two different measures of liquidity, namely the price impact and the bid-ask spread, we trained two distinct linear regressions with each measure as the dependent variable. This allowed us to investigate which one yielded a better result. The independent variables, namely "High," "Low," "Close," "Open," "Volume," and "Tether_Volume," will remain the same for all the duration of the study.

Thus, after splitting the dataset into the train and test sets and running the linear regression with the price impact as the dependent variable, we immediately realized we couldn't pursue further analysis. In fact, the linear regression model resulted in an R-squared value of 0.0270 for the training set and an R-squared value of 0.0049 for the test set, indicating that the price impact as the dependent variable wasn't suitable for explaining the relationship with the independent variables. Nevertheless, the linear regression model with the bid-ask spread as the dependent variable produced a positive R-squared of 0.7868 for the training set and of 0.7175 for the test set, which instead explains a significant portion of its relationship with the independent variables of the model. Given the limitation of the price impact variable, it is reasonable to proceed with the analysis with the bid-ask spread as the dependent variable. Therefore, we selected the second model for the continuation of our analysis.

To infer a positive impact on liquidity, it is necessary to analyze the coefficient of the variable proxy for tokenization in the regression model, in this case represented by the volume of the tokenized assets. As evidenced by Figure 5, the coefficient of "Tether_Volume" is both positive and statistically significant, thus indicating a positive relationship between the independent and the dependent variables and suggesting that an increase in trading volume is associated with enhanced liquidity.

*Figure 5: Coefficient and significance of the independent variables*

| Variable | Coefficient | Statistically Significant |
|---|---|---|
| High | 4.11604e-69 | True |
| Low | 2.27947e-95 | True |
| Close | 0.152329 | False |
| Open | 0.00743737 | True |
| Volume | 4.29033e-83 | True |
| Tether_Volume | 6.39727e-09 | True |

These findings confirm our research hypothesis, showing that tokenization (represented by our proxy variable "Tether_Volume") has a positive impact on liquidity. Holding all other factors constant, for each unit increase in tokenization, the dependent variable, given by the liquidity metric, tends to improve by the value of the coefficient.

Figure 6 displays ANOVA results for both the training set and the test set. It is shown that the F-statistic for the variable "Tether_Volume" is relatively high even though not the highest, likely because the feature comes from a different dataset. The p-values are very close to zero in both tables. This further verifies the fact that "Tether_Volume" is a significant predictor of the dependent variable.

*Figure 6: ANOVA test results for the training set (top) and test set (bottom)*

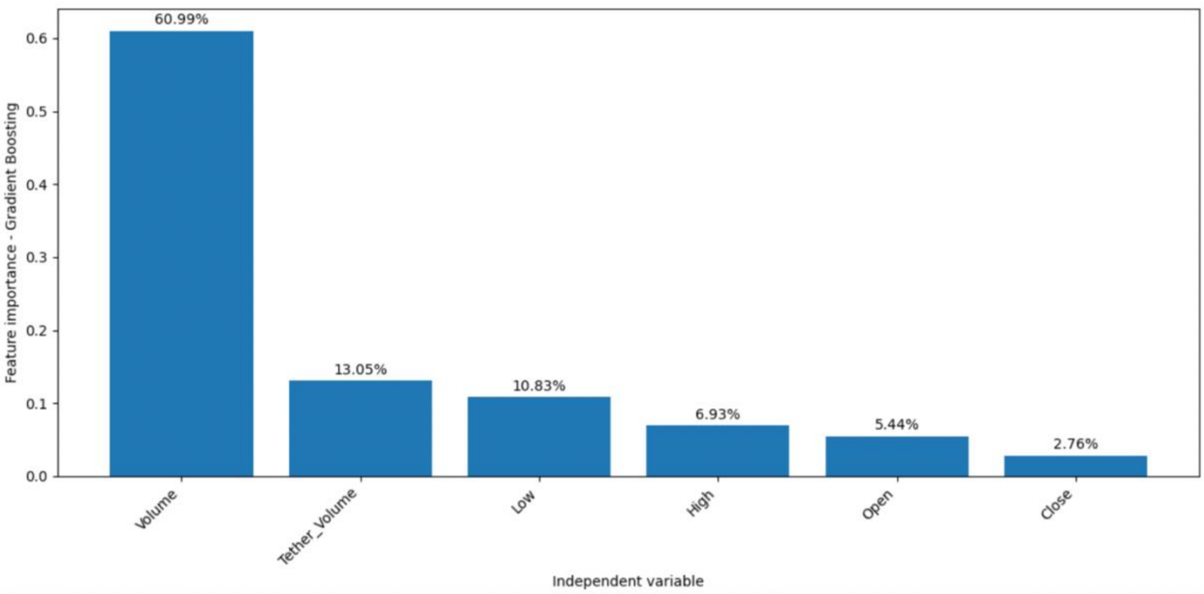
| Variable | DF | Sum Sq | F | PR(>F) |
|---|---|---|---|---|
| High | 1 | 44.7164 | 178.803 | 1.30301e-35 |
| Low | 1 | 69.1753 | 276.604 | 8.07529e-51 |
| Close | 1 | 0.623897 | 2.49472 | 0.114791 |
| Open | 1 | 0.529577 | 2.11757 | 0.146175 |
| Volume | 1 | 67.5668 | 270.172 | 7.09079e-50 |
| Tether_Volume | 1 | 2.8436 | 11.3704 | 0.00079745 |

| Variable | DF | Sum Sq | F | PR(>F) |
|---|---|---|---|---|
| High | 1 | 53.1459 | 180.453 | 6.82419e-36 |
| Low | 1 | 72.26 | 245.354 | 3.54189e-46 |
| Close | 1 | 0.0838531 | 0.284717 | 0.593836 |
| Open | 1 | 1.31795 | 4.475 | 0.0348319 |
| Volume | 1 | 52.8974 | 179.609 | 9.41285e-36 |
| Tether_Volume | 1 | 6.91192 | 23.4689 | 1.64252e-06 |

Finally, to gain further insights about the importance of the features within our dataset, we decided to analyze the performance of the gradient boosting and the random forest models.

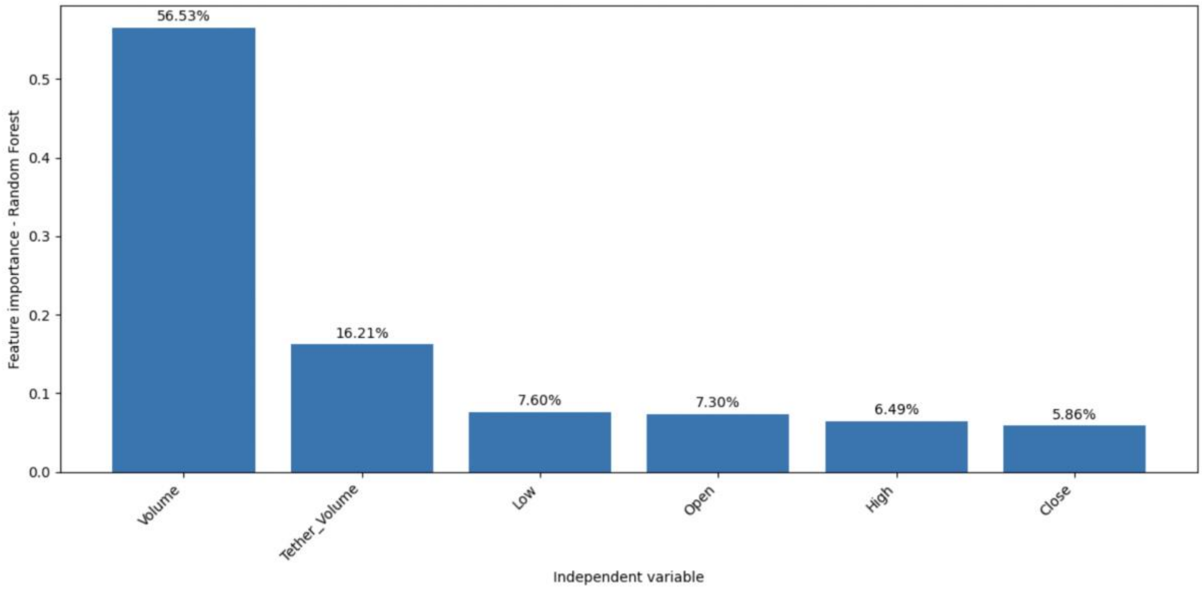
The gradient boosting model achieved an R-squared value of 0.8274 for the training set and of 0.5846 for the test set. The R-squared value for the test set is lower than the one we got through the linear regression model, but still considered high enough to say that the model provides a good fit of the data. In the following Figure 7, we observe the importance of each of the independent variables evaluated in the model. Higher values indicate more important features, and we observe that the two most important ones are "Volume" and "Tether_Volume", in this order. Even though "Volume" has a higher role in determining liquidity in this model, "Tether_Volume" adds to the predictive power of the model, suggesting that tokenization also has a meaningful positive impact on liquidity. The variable "Volume" might be more significant because the proxy of tokenization has been merged into the original dataset. Having said that, the fact that "Tether_Volume" is the second most significant feature, despite coming from a different dataset, continues to demonstrate that the tokenization of securities would help increase their liquidity.

*Figure 7: Feature importance of the model through the Gradient Boosting model*




This is even more evident when estimated through a random forest regression. In Figure 8, we observe the improved results.

*Figure 8: Feature importance of the model through the Random Forest model*




In fact, the table above reveals that "Tether_Volume" plays a significant role again, exceeding 16% of the overall feature importance of the independent variables. This further proves our research hypothesis that tokenization contributes to improving liquidity within the financial market.

Finally, we want to compare the linear regression, the gradient boosting regression and the random forest regression models to assess which one performs best. In Figure 9, we observe that the random forest model achieves the highest R-squared and the lowest Mean Squared Error (MSE) for the training set, thus standing out as the best model out of the three. The high R-squared indicates that the dependent variable explains a meaningful portion of its relationship with the independent variables in the model. The very low, almost close to 0 MSE suggests that the random forest's predictions have smaller errors compared to the other two models.

*Figure 9: Comparison between the models*

| | Model | R-squared Training set | R-squared Test set | MSE Training set | MSE Test set |
|---|---|---|---|---|---|
| **0** | Linear Regression | 0.786829 | 0.717546 | 0.247016 | 0.303162 |
| **1** | Gradient Boosting | 0.827453 | 0.584635 | 0.202512 | 0.367682 |
| **2** | Random Forest | 0.940245 | 0.540811 | 0.070132 | 0.406475 |

However, when evaluating the results regarding the test set, the random forest model performs worse, and the linear regression model seems to be the best model both in terms of R-squared and MSE. Nevertheless, this could likely be due to overfitting. To address this issue and determine if this is the case, we can apply feature selection, keeping only the most important features in the model. By discarding the variables with lower percentages of feature importance such as “Close,” “High,” and “Open,” the results for both R-squared and MSE improve, showing that overfitting might in fact be affecting the performance of the random forest model on the test set.

*Figure 10: Comparison between the models, with reduced overfitting*

| | Model | R-squared Training set | R-squared Test set | MSE Training set | MSE Test set |
|---|---|---|---|---|---|
| 0 | Linear Regression | 0.501201 | 0.522800 | 0.557217 | 0.486978 |
| 1 | Gradient Boosting | 0.759369 | 0.524026 | 0.268813 | 0.485727 |
| 2 | Random Forest | 0.936309 | 0.641840 | 0.071150 | 0.365499 |

# 4. Conclusions

Finance and technology will continue to evolve together. Technology has the power to reshape the financial landscape, and the tokenization of financial instruments serves just as one example of its impact.

This paper represented an attempt to apply quantitative data to study the effect that the potential next generation of securities will have on financial markets. It investigated the impressive benefits tokenization would bring while shedding light on its risks and threats. It unveiled the lack of a consistent regulatory framework and illustrated the present legal landscape. Even though the shift to DLT infrastructures and blockchain is not expected to happen in the short-term and is more likely to be deployed in a progressive manner, the tokenization of financial instruments remains an extremely promising emerging field. Especially when talking about securities, the related literature and empirical analyses are emerging at the moment, but the continuously growing interest in the field, both from big corporations and single individuals, demonstrates how exciting future studies and implementations of the tokenization process prospect to be.

By analyzing some of the existing literature on the tokenization applied to assets, particularly focusing on Stablecoins, it is found that cash tokens have a liquidity ratio that is at least double the requirements of typical banks. In fact, the Liquidity Coverage Ratio (LCR), computed as the ratio of High-Quality Liquid Assets (HQLA) to the outflow over the following 30 days, averaged over 200%. This represents a significant improvement compared to the 118% LCR computed for eight U.S. Systemically Important Banks in 2022. Following the premise that tokenization could significantly improve liquidity, our empirical analysis presented a liquidity metric, the bid-ask spread, as the dependent variable of our models. Since this measure wasn't readily available in the chosen dataset, we feature engineered it through a formula validated by previous research. The first model that was implemented, the linear regression, showed a positive and statistically significant coefficient, thus confirming our research hypothesis claiming that more tokenization enhances liquidity. To further endorse this finding, we employed a gradient boosting regression and a random forest regression to study the feature importance of the independent variables. The variable proxy for the tokenization, "Tether_Volume", reached 16% of the explanation of the model through the random forest regression. While it is true that the importance of the "Tether_Volume" value is lower than the actual "Volume" variable, this is likely because the proxy for tokenization sources from a different dataset. Despite all this, the value "Tether_Volume" remains the second largest feature in importance in our analysis. This further supports our research hypothesis, suggesting that the tokenization of financial instruments increases the model's predictive power, thus suggesting, once again, that tokenization has a positive impact on liquidity.

There are a few limitations to both the theoretical side and the practical side of the analysis. Starting from the theoretical side, the claim that tokens could democratize finance and offer investment opportunities to individuals even in underserved regions is not a plausible reality everywhere, especially in countries with unstable currencies such as the African states. Moreover, even though the potential of increasing liquidity within the market is indeed very high, it represents a double-edged weapon. In fact, it can have great outcomes on highly illiquid assets and securities, but on the other side, it presents the risk of drying up liquidity in off-chain markets, putting forward the threat of arbitrage. The business rationale for the adoption of the tokenization process in markets with limited liquidity and considerable disintermediation is very likely. Nevertheless, its overall usage might be more difficult to justify in markets consisting of predominantly developed economies. This is also because the cost of enhancing the DLT infrastructure in such markets might be more elevated than the benefits.

Regarding the empirical side of the paper, it stands to reason that a more accurate analysis would require additional data that in this case was feature engineered, specifically the feature regarding bid-ask spreads. This other approach would allow for a more thorough study of liquidity. Furthermore, our analysis is limited by the lack of a dataset on tokenized securities, forcing us to merge two datasets and use a proxy for tokenization.

Therefore, as an objective for future research, it is suggested to focus on compiling a comprehensive dataset specifically containing information on tokenized securities. In fact, a direct analysis of the impact that tokenization has on liquidity metrics would yield more reliable results. When assembling the dataset on tokenized securities, it is also recommended to analyze a longer time-period by incorporating more recent data to delve deeper into the current effects of tokenization on the liquidity of the financial markets. Finally, to consider the differences across the different market types mentioned above (developed and developing economies), a comparative study would help in assessing the varying impacts of tokenization, together with the underlying factors at play.

To conclude, the findings of this research suggest that the tokenization process of financial instruments has all the potential to drastically reshape the current financial landscape by enhancing liquidity, as we demonstrated, and by providing brand new opportunities for investment. Despite the challenges and risks associated with tokenization, the benefits still highly seem to exceed the drawbacks. Therefore, as technology continues to improve, the role of tokenization is expected to become more and more significant. There is much hope for the exciting developments that the process of tokenization will bring: individuals, professionals and big corporations all seem eager to delve deeper into its immense potential.

## 5. References


Abdi, F., & Ranaldo, A. (2017, August 26). *A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices*. OUP Academic. https://academic.oup.com/rfs/article/30/12/4437/4047344

Adegboye, A. (2023, July 12). *"Crypto could revolutionize finance" BlackRock CEO*. Mariblock. https://www.mariblock.com/crypto-could-revolutionize-finance-blackrock-ceo/#:~:text=Validation%20from%20influential%20people%20like

Arner, D. W., Auer, R., & Frost, J. (2021, December 15). *Stablecoins: Risks, Potential and Regulation*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3979495

Benedetti, H., & Rodriguez-Garnica, G. (2022, May 10). *Tokenized Assets and Securities*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4069119

Birch, D. G. W. (2023, March 1). *Larry Fink Says Tokens are "The Next Generation for Markets."* Forbes. https://www.forbes.com/sites/davidbirch/2023/03/01/larry-fink-says-tokens-are-the-next-generation-for-markets/?sh=711d68447bb1

Bouchaud, J. P. (2009, March 13). *Price Impact*. ResearchGate. https://www.researchgate.net/publication/46462472_Price_Impact

Carapella, Francesca, Grace Chuan, Jacob Gerszten, Chelsea Hunter, and Nathan Swem (2023). *Tokenization: Overview and Financial Stability Implications*. Finance and Economics Discussion Series 2023-060. Washington: Board of Governors of the Federal Reserve System. https://doi.org/10.17016/FEDS.2023.060

Cohen, R., Smith, P., Arulchandran, V., & Sehra, A. (2018, March). *Automation and blockchain in securities issuances*. Allen & Overy. www.allenovery.com

Dahlborn, A., de la Roche, M., Kajtazi, L., & Daykin, R. (2023, March 22). *Tokenization of Assets and Blockchain*. EUBlockchain. https://www.eublockchainforum.eu/sites/default/files/research-paper/Tokenization of Assets & Blockchain_0.pdf

Falempin, L., Coheur, D., Van Hecke, P., & Walsh, E. (2019). *Tokenized Securities*. Tokeny. https://tokeny.com/wp-content/uploads/2019/01/TOKENIZED-SECURITIES.pdf

Heines, R., Dick, C., Pohle, C., & Jung, R. (2021). *The tokenization of everything: Towards a framework for understanding the potentials of tokenized assets*. AIS Electronic Library (AISeL). https://aisel.aisnet.org/pacis2021/40/

Hussain, M. B. (2023, May 22). *Nvidia Stocks Historical Data*. Kaggle. https://www.kaggle.com/datasets/bilalwaseer/nvidia-stocks-historical-data

Liao, G. (2022, September 29). *Macroprudential Considerations for Tokenized Cash*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4228268

Liao, G. (2023, March 25). *Payment versus trading stablecoins*. CEPR. https://cepr.org/voxeu/columns/payment-versus-trading-stablecoins#:~:text=Insufficiently%20backed%20stablecoins%20risk%20runs,deposits%20to%20the%20banking%20system

Liao, G. Y., & Caramichael, J. (2022, January). *Stablecoins: Growth Potential and Im- pact on Banking*. International Finance Discussion Papers 1334. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/IFDP.2022.1334

McKinsey & Company. (2023, October 6). *What is tokenization?*. McKinsey & Company. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-tokenization

Misura, P., & Wirth, M. (n.d.). *How tokenization of assets is shaping the Financial Landscape: A comprehensive analysis*. Horn Company. https://www.horn-company.de/blog-en/crypto-en/how-tokenization-of-assets-is-shaping-the-financial-landscape/?lang=en

Muoka, R. (2023, October 17). *Unlocking a $16 Trillion Opportunity: The future of Asset Tokenization*. LinkedIn. https://www.linkedin.com/pulse/unlocking-16-trillion-opportunity-future-asset-robert-muoka/#:~:text=In%20March%202023%2C%20Larry%20Fink,and%20real%2Dworld%20assets.%22

OECD (2020), *The Tokenisation of Assets and Potential Implications for Financial Markets, OECD Blockchain Policy Series*, www.oecd.org/finance/The-Tokenisation-of-Assets-and-Potential- Implications-for-Financial-Markets.htm

*President's Working Group on Financial Markets Releases Report and Recommendations on Stablecoins*. U.S. Department of the Treasury. (2021, November 1). https://home.treasury.gov/news/press-releases/jy0454

ProgrammerRDAI. (2022, May 24). *Tether Crypto Price*. Kaggle. https://www.kaggle.com/datasets/ranugadisansagamage/tether-crypto-price

World Bank Group. (2023, August 14). *The global findex database 2021*. World Bank. https://www.worldbank.org/en/publication/globalfindex