Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUSA

Florence, Italy
11- 13 September 2025

AMASES 2025

Special session: Decentralized Exchanges and Tokenization

MUSA

TOKENIZATION AND PRODUCTION THEORY: INNOVATION, EFFICIENCY, AND MARKET DYNAMICS

Leonardo Maria De Rossi, Huynh Thanh Thien Pham, Claudio Tebaldi
Università Bocconi

The authors thank Michele Russo for excellent research assistance and gratefully acknowledge funding from SDA Bocconi DIR project 'Tokenization'. C.T. and H.P. acknowledge funding from MUR-Prin 2022- Prot. 20227TCX5W, Next Generation EU.

MUSA

MULTILAYERED URBAN SUSTAINABILITY ACTION

MUSA – Multilayered Urban Sustainability Action is an Innovation Ecosystem funded by the Ministry of University and Research under the National Recovery and Resilience Plan.

Context

Web Revolution

The evolution of **the Internet** has been changing our life. From the introduce of static websites to the emerging web platforms enabling users with even more power in the digital world.

Web3

Web3 disrupts by embedding ownership and incentives into platforms through **tokens** and **decentralization**.

Digital Tokens

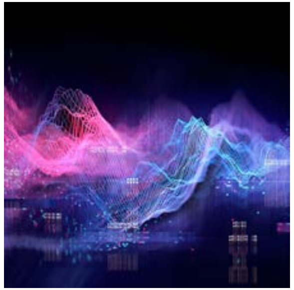
ERC20 utility tokens are digital access keys that enable transactions, services, and governance on blockchain-based platforms.

Digital Platform

Web2 disrupted industries by turning digital platforms into engines of scale through **network effects** and **user data**.

Blockchain

Serving as the backbone technology of the evolution, blockchain enables users to **transact directly with trust.** At its heart, **smart contract** is a powerful engine.



Research Motivation

While tokens are being traded more and more, the understanding of its **fundamental drivers** is still underexplored.

Without such understanding value of these tokens are nearly just **speculation**.

<u>Seeks fundamentals for pricing utility token.</u>

Research Questions

What drives the value of ERC20 utility tokens?

What data was collected?

What analysis has been done on the collected data?

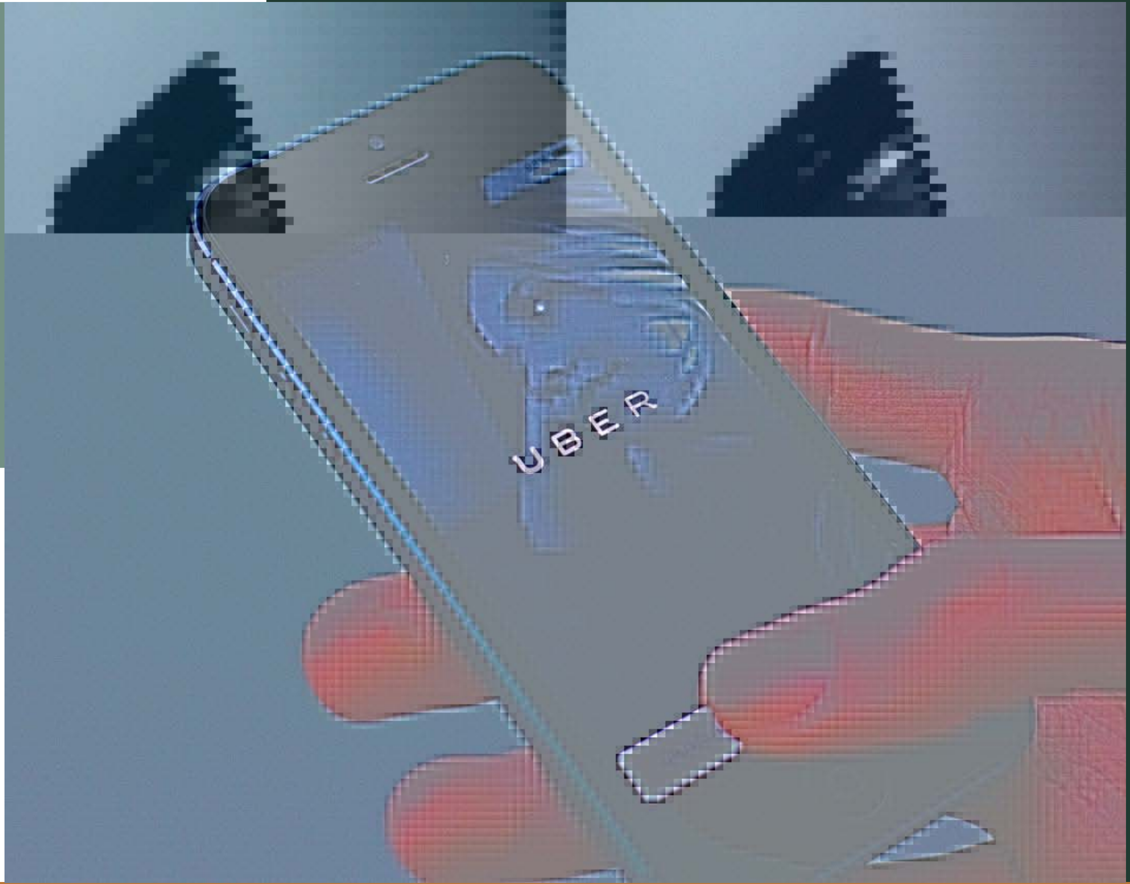
How empirical data can be linked to theoretical valuation models?

Inspiring Analogy

Owning a token is similar to joining a digital platform.

Success is the ability to generate **network effects**.

=> Parallel to work of Rochet and Tirole (2006) on two-sided markets.



Dataset Overview

~4k	>30	8
Tokens	Variables	Years

Descriptive Statistics

Variables span **market**, **technical**, and **social** dimensions.

Many **time-variant** variables.

Huge **heterogeneity**.

The ERC20 utility token ecosystem is extremely diverse.

Variable	Mean	Std	Min	Max
prices	1,509,426.08	259,019,615.69	0	45,019,652,550.13
market_caps	156,031,932.31	1,013,796,371.79	0	35,573,266,954.87
total_volumes	19,577,198.80	252,582,051.91	0	25,290,764,281.58
holders_count	21,727.81	68,480.79	0	846,658
total_supply	1.5607E+53	2.73481E+55	0.2941303	4.79209E+57
transfer_count	207.18	1,048.53	1	73,175
unique_senders	67.47	331.42	1	30,203
unique_receivers	84.89	533.21	1	49,870
total_sc_calls	376.38	1,775.11	1	122,008
gh_commits_count	5.39	15.22	0	275
gh_unique_committers_count	1.21	3.78	0	72
gh_prs_count	7.35	18.97	0	163
twitter_followers	70,786.45	164,614.18	0	2,055,884
telegram_users	8,068.78	15,406.03	2	261,956
platforms_count	2.28	2.05	0	15
exchanges_count	13.09	20.39	0	85
exchangeable_pairs_count	6.98	9.82	0	98

Data Collection

Python scripts

API calls (rate limit!!!)

This code is only for GitHub data.

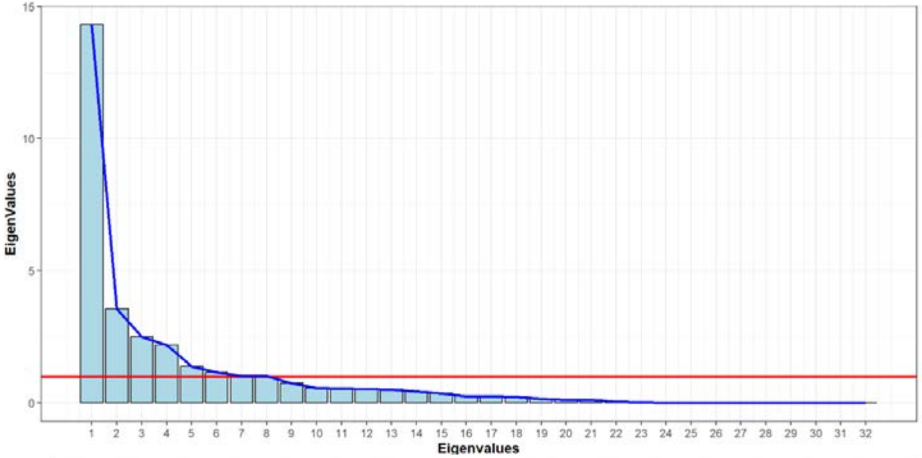
PCA pre-processing:

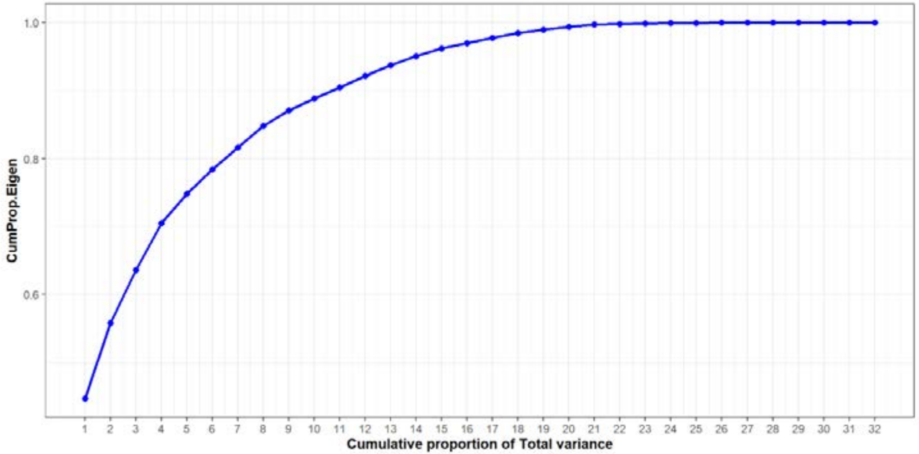
- Filtering out tokens no longer traded.
- Removing sparsely populated variables.
- Excluding extreme outliers (just for PCA).
- Log-transformed to avoid skewness.

```python
def get_gh_commits(gh_path, date_from, date_till):
    url = f"https://api.github.com/repos/{gh_path}/commits?since={date_from}&until={date_till}"
    headers = {
        "Accept": "application/vnd.github.v3+json",
        "Authorization": "Bearer <api_key>",
        "X-GitHub-Api-Version": "2022-11-28"
    }
    response = requests.get(url, headers=headers)
    return response.json()
def build_gh_time_series(paths, first_date='01/01/2001', last_date='15/12/2024'):
    start_date = pd.to_datetime(first_date, format='%d/%m/%Y').strftime('%Y-%m-%d')
    end_date = pd.to_datetime(last_date, format='%d/%m/%Y').strftime('%Y-%m-%d')
    date_range = pd.date_range(start=start_date, end=end_date, freq='ME')
    api_dates = [date.strftime('%Y-%m-%d') for date in date_range]
    for date in api_dates: # Get actitivies data for each date in the time series
        from_date = pd.to_datetime(date).replace(day=1)
        for path in paths: # Because one project might have multiple repos, so loop thru them
            commits_data = get_gh_commits(path, from_date.strftime('%Y-%m-%d'), date)
            time.sleep(1.5)
            prs_data = get_gh_prs(path, from_date.strftime('%Y-%m-%d'), date)
            time.sleep(1.5)
            issues_data = get_gh_issues(path, from_date.strftime('%Y-%m-%d'), date)
            time.sleep(1.5)
            subscribers_data = get_gh_subscribers(path, from_date.strftime('%Y-%m-%d'), date)
            time.sleep(1.5)
    return df
```

Cumulative proportion of expl. variances

	prop_1:1	prop_1:2	prop_1:3	prop_1:4	prop_1:5	prop_1:6	prop_1:7	prop_1:8	prop_1:9	prop_1:10
total_volumes_mean	0.66	0.69	0.71	0.85	0.88	0.89	0.89	0.89	0.89	0.89
holders_count_mean	0.27	0.28	0.29	0.31	0.42	0.43	0.47	0.47	0.85	0.85
total_supply_mean	0	0.13	0.78	0.95	0.95	0.95	0.95	0.95	0.95	0.95
total_top_holding_mean	0	0.12	0.74	0.91	0.91	0.91	0.91	0.91	0.91	0.91
top_holders_count_mean	0	0	0	0	0	0.24	0.76	0.91	0.97	1
total_amount_transfered_mean	0.01	0.1	0.57	0.67	0.68	0.68	0.68	0.68	0.68	0.68
transfer_count_mean	0.96	0.96	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98
unique_senders_mean	0.93	0.94	0.94	0.94	0.96	0.96	0.96	0.96	0.97	0.97
unique_receivers_mean	0.88	0.9	0.9	0.9	0.93	0.94	0.94	0.94	0.94	0.94
total_sc_calls_mean	0.77	0.78	0.82	0.97	0.98	0.98	0.98	0.98	0.98	0.98
succeeded_sc_call_mean	0.77	0.77	0.82	0.97	0.98	0.98	0.98	0.98	0.98	0.98
external_sc_calls_mean	0.87	0.89	0.89	0.93	0.97	0.97	0.97	0.97	0.98	0.98
internal_sc_calls_mean	0.62	0.63	0.7	0.94	0.97	0.97	0.97	0.97	0.97	0.98
sc_caller_count_mean	0.93	0.94	0.94	0.94	0.96	0.96	0.96	0.96	0.97	0.97
transactions_made_mean	0.96	0.96	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98
transfer_total_calls_mean	0.86	0.89	0.89	0.9	0.9	0.9	0.92	0.92	0.95	0.95
transfer_internal_count_mean	0.67	0.67	0.69	0.7	0.72	0.72	0.75	0.76	0.82	0.82
transfer_external_count_mean	0.76	0.79	0.81	0.92	0.96	0.96	0.96	0.97	0.97	0.97
transferFrom_total_calls_mean	0.72	0.72	0.76	0.87	0.87	0.87	0.9	0.9	0.94	0.94
transferFrom_internal_count_mean	0.7	0.7	0.74	0.86	0.87	0.87	0.9	0.9	0.94	0.94
transferFrom_external_count_mean	0.53	0.56	0.58	0.7	0.73	0.73	0.73	0.74	0.74	0.75
gh_commits_count_mean	0.04	0.56	0.57	0.62	0.63	0.77	0.77	0.91	0.91	0.91
gh_unique_committers_count_mean	0.02	0.45	0.47	0.5	0.51	0.69	0.69	0.91	0.91	0.91
gh_prs_count_mean	0.06	0.57	0.64	0.66	0.67	0.72	0.72	0.8	0.8	0.84
gh_issues_count_mean	0.05	0.51	0.58	0.58	0.62	0.62	0.63	0.68	0.68	0.91
gh_stargazers_count_mean	0.09	0.67	0.71	0.71	0.71	0.73	0.73	0.75	0.76	0.85
gh_subscribers_count_mean	0.08	0.56	0.62	0.64	0.64	0.7	0.7	0.74	0.75	0.87
twitter_followers_mean	0.24	0.24	0.26	0.37	0.44	0.52	0.64	0.68	0.68	0.69
platforms_count	0.09	0.1	0.12	0.22	0.54	0.59	0.67	0.72	0.72	0.73
telegram_users	0.08	0.09	0.1	0.16	0.26	0.57	0.69	0.79	0.81	0.82
exchanges_count	0.36	0.36	0.42	0.54	0.66	0.66	0.66	0.69	0.7	0.71
exchangeable_pairs_count	0.33	0.34	0.36	0.37	0.66	0.68	0.69	0.71	0.76	0.76





Principal Component Analysis

The performance is quite satisfactory, showing (from the plots) that we should retain 3-4 PCs for a good and extensive representation of the original set of variables

Principal Component Analysis

- Run full sample analysis.
- Confirm results with subsamples.
- Consistent results across all samples.
- The first 3 components explain more than 60% of variance.

Dimensionality reduction

A reduced set of components can effectively capture the major dimensions of variability in the dataset.

Latent variables identification

Network usage variables cluster strongly on **PC1**, **development** activity on **PC2**, and token **supply** and **top holdings** on **PC3**.

	PC1	PC2	PC3	PC4	PC5
total_volumes_mean	0.81	-0.16	0.14	-0.38	-0.17
holders_count_mean	0.52	-0.1	0.1	-0.16	-0.33
total_supply_mean	0.04	-0.36	0.81	0.41	-0.01
total_top_holding_mean	0.02	-0.35	0.79	0.41	0
top_holders_count_mean	0.02	0.02	0.06	0.03	-0.01
total_amount_transfered_mean	0.11	-0.3	0.69	0.32	0.04
transfer_count_mean	0.98	-0.08	-0.09	0.09	-0.01
unique_senders_mean	0.96	-0.09	-0.02	-0.04	-0.13
unique_receivers_mean	0.94	-0.13	0.01	-0.05	-0.19
total_sc_calls_mean	0.88	0.04	-0.22	0.38	0.11
succeeded_sc_call_mean	0.88	0.04	-0.22	0.38	0.11
external_sc_calls_mean	0.93	-0.14	0.07	-0.2	-0.2
internal_sc_calls_mean	0.79	0.08	-0.27	0.49	0.18
sc_caller_count_mean	0.96	-0.09	-0.01	-0.05	-0.13
transactions_made_mean	0.98	-0.05	-0.08	0.08	-0.06
transfer_total_calls_mean	0.93	-0.15	-0.01	-0.11	-0.03
transfer_internal_count_mean	0.82	-0.09	-0.13	0.11	0.13
transfer_external_count_mean	0.87	-0.18	0.13	-0.33	-0.21
transferFrom_total_calls_mean	0.85	0.06	-0.19	0.33	0.05
transferFrom_internal_count_mean	0.83	0.07	-0.2	0.35	0.05
transferFrom_external_count_mean	0.73	-0.18	0.12	-0.34	-0.19
gh_commits_count_mean	0.2	0.72	0.11	0.24	-0.1
gh_unique_committers_count_mean	0.15	0.66	0.13	0.17	-0.08
gh_prs_count_mean	0.25	0.71	0.26	-0.11	-0.1
gh_issues_count_mean	0.22	0.68	0.26	-0.06	-0.19
gh_stargazers_count_mean	0.3	0.76	0.21	0	-0.02
gh_subscribers_count_mean	0.28	0.69	0.26	-0.13	0.03
twitter_followers_mean	0.49	0.02	0.15	-0.32	0.28
platforms_count	0.29	0.12	0.15	-0.31	0.57
telegram_users	0.28	-0.09	0.1	-0.25	0.32
exchanges_count	0.6	0.03	0.24	-0.35	0.35
exchangeable_pairs_count	0.58	0.08	0.13	-0.1	0.54

Interpretation of PCA Components

Across all years, the first three principal components consistently explain more than 60% of the variance, and the grouping of variables remains stable.

01

Network Activity

This component quantifies high-frequency transactional activity and user participation, capturing activity such as transfers, volumes, and smart contract calls.

02.

Developer Activity

A development and productivity factor that highlights the **production side** of the emergent economy, tracked by GitHub activity and coding intensity.

03.

Supply & Concentration

A factor that reflects token supply and distribution dynamics, capturing a structural dimension of the token ecosystem.

From Empirical Results to Theoretical Models

How PCA results map to theoretical findings?

Theoretical Baseline: Cong, Li, Wang (2021)

- L. W. Cong, Y. Li, and N. Wang, "Tokenomics: Dynamic Adoption and Valuation", *The Review of Financial Studies*, vol. 34, no. 3, 2021.
- The authors wanted to shed light on the argument:"*Tokens help grow the ecosystem and allow all participants to benefit from the growth prospect of platforms.*"
- Deriving a PDE that is similar to Black-Scholes formula.
- Solving for an equilibrium pricing equation grounded in network effects and productivity shocks.
- Rigorous economic foundation for token valuation.

The Gordon Growth formula

Classical equity valuation formula:

$$P_t = \frac{D_{t+1}}{(r - g)} \qquad (1)$$

where:

- g : Constant growth rate expected for dividends.
- r : Interest rate.
- D_{t+1} : Value of next dividends.

Observable dividends are the key value driver.

Token Pricing vs Derivative Pricing (Black-Scholes formula)

The PDE for the token price $P_t = P(A_t)$ in Cong, Li, Wang (2021):

$$\mu_t^P P_t = \mu^A A_t \frac{\partial P_t}{\partial A_t} + \frac{1}{2}\left(\sigma_t^A A_t\right)^2 \frac{\partial^2 P_t}{\partial {A_t}^2} \qquad (2)$$

Underlying A_t (productivity, non-tradable) | P_t: token price | μ^A, σ^A: drift/vol of productivity

The Black-Scholes formula with **no-arbitrage** condition:

$$rV = rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}$$

Underlying S (tradable) | V: derivative price | σ: volatility | r: interest rate

Utility vs Carry Cost Trade-off

With tokens, agent i's decision on the real balance of medium of exchange, the effective carry cost is now $r - \mu_t^P$, and **the optimality condition agents' token balance** is where the marginal benefit of transaction is equal to the carry cost:

$$r - \mu^P(A_t) = (1 - \alpha)\left(\frac{N_t A_t e^{u_i}}{x_{i,t}^*}\right)^{\alpha} > 0 \qquad (3)$$

Dynamic Token Valuation Model

Equilibrium token pricing formula in Cong, Li, Wang (2021):

$$P(A_t) = \left(\frac{N(A_t)\, S(A_t)\, A_t}{M}\right)\left(\frac{1-\alpha}{r - \mu^P(A_t)}\right)^{\frac{1}{\alpha}} \qquad (4)$$

- Extension of Gordon Growth to tokens.
- Cash flows unobservable → replaced by productivity process.

Dynamic Token Valuation Model

The variables in the equilibrium:

$$P(A_t) = \left(\frac{\underbrace{N(A_t)}_{\text{User base}} \times \underbrace{S(A_t)}_{\text{Transaction needs}} \times \underbrace{A_t}_{\text{Productivity}}}{\underbrace{M}_{\text{Token supply}}} \right) \left(\frac{1-\alpha}{r - \mu^P(A_t)} \right)^{\frac{1}{\alpha}}$$

Mapping Empirical Components on the Theoretical Model

The latent variables identified by the PCA can be mapped nicely onto the token pricing equilibrium proposed in Cong, Li, Wang (2021):

Empirical Components	Theoretical Variables
PC1. Network Activity	$N(A_t)$ = User base *(daily active addresses)*
PC1. Network Activity	$S(A_t)$ = Transactional needs *(daily trading volume)*
PC2. Developer Activity	A_t **= Productivity**
PC3. Supply and Concentration	M = Token supply

Policy & Market Implications

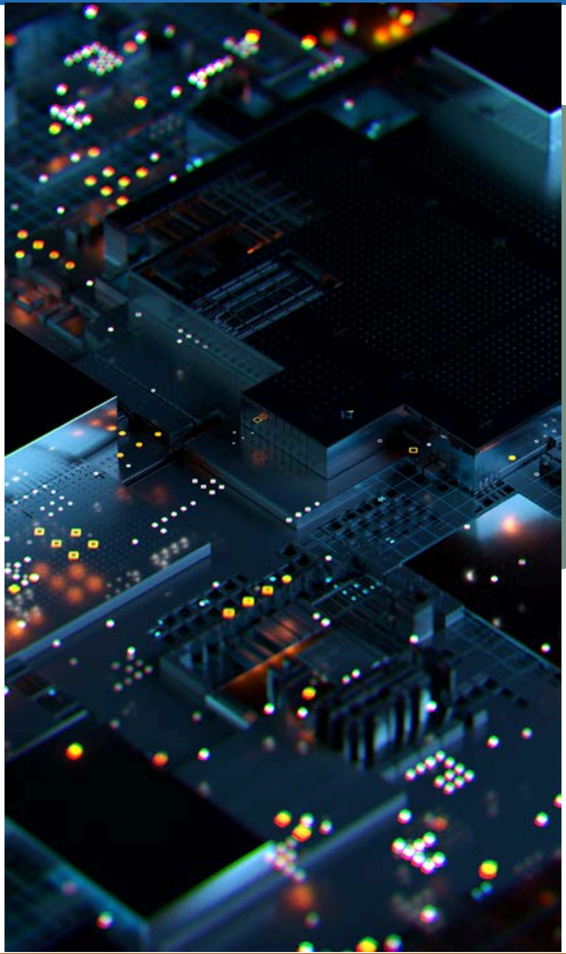
The findings also have several broader implications for token design, investment strategies, regulation, and draw the connection between empirical data and theoretical models.

Token design: Development might play an important role in observing the productivity of the token which implies to be an influential information on token price.

Investor strategy: Fundamentals beyond volume and conventional financial data. Our preliminary results attest non-financial factors as the latent drivers instead of financial ones.

Regulation: Disclosure of system usage and development metrics should be encouraged to attract fundamental investors.

Theoretical link: empirically identified latent drivers provide preliminary proofs to the theory on "new fundamentals" for token pricing.



Conclusion

First empirical link between Cong, Li, Wang (2021)'s model and ERC20 tokens.

- ERC20 token pricing resembles derivative valuation more than equity.
- Productivity and adoption matter as key hidden drivers.
- Supply concentration plays an important but complex role.

Next steps:

- Regression analysis, robustness checks, sub-sample analysis.
- Extend framework to NFTs and real-world asset tokens.

Provisional Next Step: Regression Analysis

Testable empirical implication, from eq. 4:

$$H_0:\ ln\frac{P_{i,t+1}}{P_{i,t}} = \Delta\, ln(N_{i,t}\, S_{i,t}) + \Delta\, ln\, A_{i,t} - \Delta\, ln\, M_{i,t} - \frac{1}{\alpha}\,\Delta\, ln\left(r_t - \mu_t^P\right)$$

$$ln\frac{P_{i,t+1}}{P_{i,t}} = \mu_i + \beta_1\,\Delta\, ln(N_{i,t}\, S_{i,t}) + \beta_2\Delta\, ln\, A_{i,t} + \beta_3\Delta\, ln\, M_{i,t} + \varepsilon_{i,t+1} + Control\ Variables$$

By taking logs, we have returns on the left-hand side.
This allows us to run regressions of returns on changes in our PCA factors.
Econometric test: PC factors significantly explain returns?

Note: $\frac{1}{\alpha}\,\Delta\, ln\left(r_t - \mu_t^P\right)$ *is dropped in regression because it is global effective carry cost.*

Provisional Next Step: Potential Regressors

Name	Variable	Interpretation
Dependent Variable	$ln\frac{P_{i,t+1}}{P_{i,t}}$	Observed token price changes
Regressor 1– $\boldsymbol{PC1_{i,t}}$	$\Delta\, ln(N_{i,t}\, S_{i,t})$	Proxy for **network effects** (holders, transfers, active users)
Regressor 2 - $\boldsymbol{PC2_{i,t}}$	$\Delta\, ln\, A_{i,t}$	Proxy for **platform productivity** (GitHub activity, commits, issues)
Regressor 3 - $\boldsymbol{PC3_{i,t}}$	$\Delta\, ln\, M_{i,t}$	Proxy for **structural token supply** (total supply and distribution)

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

MUSA

MUSA

Thank you for listening

Huynh Thanh Thien Pham
huynh.pham@unibocconi.it

MUSA Multilayered Urban Sustainability Action