Is the Market Blind to Data? An Empirical Study of Valuation Multiples in Latvia

Ritvars Mētra
ORCID iD: 0009-0001-1102-4353
sz19002@lbtu.lv
Latvia University of Life Sciences and Technologies, Faculty of Economics and Social Development
Jelgava, Latvia

https://doi.org/10.18267/pr.2026.vol.2587.23

Abstract: This paper examines whether capital markets reward firms that treat data as a central economic asset. Drawing on the concept of Data Capitalism, it tests the existence of a "data premium" in firm valuation using a novel Data Capitalism Maturity Framework applied to Latvia's 101 most valuable firms. Combining a 2024 cross-sectional analysis with panel data covering 2022–2024, the study evaluates whether higher data maturity is associated with higher valuation multiples once firm size and industry are controlled for. Descriptive patterns, non-parametric correlation tests, and multivariate regressions consistently fail to identify a positive relationship between data maturity and market valuation. Instead, firm size and sectoral affiliation dominate valuation outcomes. A robust and persistent negative relationship between turnover and valuation multiples reveals a "revenue penalty," suggesting that scale is discounted rather than rewarded in this market. Industries anchored in tangible assets, such as natural resources and real estate, command higher baseline valuations, while data-intensive business models receive no systematic premium. The findings suggest that, in small and shallow capital markets, the financial logic of Data Capitalism does not translate automatically into valuation signals. Data may shape internal performance, but it remains largely invisible in market pricing.

Keywords: Data Capitalism, Firm Valuation, Valuation Multiples

JEL Classification codes: G32, L25, O33

INTRODUCTION

In the midst of ICT advancement and ubiquitous digitisation, the global economy is rearranging itself around data. This process has given birth to a new economic logic where data is not an accessory, not an exhaust of digital activity, but something closer to capital in its own right: accumulated, refined, deployed, defended. While Zuboff and Srnicek have characterised this shift through the lenses of 'surveillance' and 'platform' capitalism respectively (Srnicek, 2017; Zuboff, 2019), a broader and more integrated economic paradigm of Data Capitalism (DC) is required to understand current market dynamics. In practice, DC looks less like a switch and more like a gradient. That gradient matters. Firms don't fall neatly into "data-driven" and "not data-driven" boxes. Some live and die by data—primary data capitalists whose revenues, strategies, and internal coordination hinge on continuous data extraction and analysis. Others operate in hybrid modes, experimenting, borrowing tools, partially reorganising workflows. At the margins sit firms where data remains incidental, supportive at best. Treating these differences as noise misses something essential: the degree of data centrality is itself an economic characteristic, one that may shape how firms are perceived long before it reshapes their income statements.

Already a decade ago some scholars observed that data-driven decision-making altered firm performance (Brynjolfsson et al., 2011; McAfee & Brynjolfsson, 2012); since then, the argument has thickened, darkened, picked up sharper edges. Datafication isn't merely a technical upgrade bolted onto existing firms. It's a reworking of how labour is organised, how markets are read, how future revenues are imagined (Bolin, 2022; Pfeiffer, 2022). Digital tools and data analytics are heralded as universal fixes, often overlooking their inherent limitations and unintended consequences (Morozov, 2013). Contemporary scholarship underscores data's role as a valuable intangible asset that significantly influences firm valuations (Birch et al., 2021; Tambe, 2014). Cheong et al. demonstrate that data-driven models boost firms' valuations by quantifying data's role in revenue generation and competitive advantages in digital economies (Cheong et al., 2023). Building on these insights, this study posits the following hypothesis: The market assigns a higher valuation multiple (Value-to-Turnover ratio) to higher-tier Data Capitalists.

This is where things start to feel slightly odd. Most empirical work on DC (including its siblings—Platform Capitalism, Surveillance Capitalism, Digital Capitalism, etc.) and Data-Driven Business Models (DDBM) overwhelmingly fixates on large economies. Meanwhile, smaller economies are treated as footnotes, if they appear at all. This makes the picture incomplete. Digitalisation doesn't politely stop at national borders, and small, open economies often adopt new organisational logics quickly, if unevenly. The Baltic states are a case in point: heavy investment in digital infrastructure, high rates of administrative digitalisation, a dense ecosystem of small and mid-sized firms navigating tight markets and limited scale. These aren't replicas of Silicon Valley. They're something messier—and potentially more revealing.

Latvia, in particular, offers an instructive setting. Capital markets are thin. Information asymmetries bite harder. Growth narratives travel fast. Under such conditions, expectations can matter as much as current performance, sometimes more. This raises a simple but underexplored question: do markets recognise data intensity itself as a source of future advantage? Or, put differently, does deeper integration into Data Capitalist logics translate into higher valuation multiples—even when revenues haven't yet caught up?

This paper steps into that gap. The blind spot, then, is not theoretical—it's empirical and geographic. We know how Data Capitalism looks at the frontier. We know far less about how markets respond to it in places where scale is limited, narratives travel quickly, and data-driven futures are priced before they fully arrive. Rather than asking whether data-heavy firms are more profitable in the abstract, it asks whether the market prices their position on the DC spectrum. Using Latvian firm-level data, the analysis proceeds in two movements: a cross-sectional snapshot for 2024, followed by a panel covering 2022–2024. Firms are classified along an ordinal DC tier, allowing for gradation rather than binary labels. Valuation is captured through the value-to-turnover ratio, a measure notoriously skewed, ill-behaved, and therefore revealing.

1 LITERATURE REVIEW

In neoclassical and financial economics, market value is typically explained by current and projected earnings, asset bases, growth prospects, and risk-adjusted discounting (Damodaran, 2025). Within this framework, valuation remains tethered to realized or at least clearly forecastable economic performance. Firms with weak current financials may command higher valuations, but only insofar as credible growth trajectories can be inferred from conventional indicators such as market expansion, technological innovation, or cost advantages (Damodaran, 2009; Nissim & Penman, 2001). The underlying assumption is that future value must be legible through existing financial signals, even if imperfectly so. The logic of DC complicates this relationship. Data-intensive firms often accumulate assets whose economic potential is temporally displaced and difficult to translate into immediate revenue streams. Data generates value not only through direct monetization, but through optionality: the capacity to enable new products, optimize processes, enter adjacent markets, or extract efficiencies that are not yet realized at the time of valuation. As a result, markets increasingly price firms on the basis of anticipated data-driven capabilities rather than solely on current performance (Cheong et al., 2023; Ker & Mazzini, 2020).

Empirical research supports this shift. Olabode et al. surveyed 360 firms in the UK to measure the relationship between Big Data Analytics Capability (BDAC) and market performance. Their empirical results show that BDAC allows firms to be more "forward-looking" in their decision-making, which the market rewards with a superior "marketplace advantage (Olabode et al., 2022). Wamba et al. provide quantitative evidence that high Big Data Analytics Capability (BDAC) is significantly and positively associated with both financial performance and market value. They argue that because the market is efficient at pricing these dynamic capabilities, the firm's valuation reflects its anticipated ability to adapt and grow in data-rich environments, which the authors identify as a primary driver of market value beyond traditional accounting metrics (Wamba et al., 2017). Lee et al. observed Statistically Significant Positive Abnormal Returns (a jump in stock price) immediately after a firm announced a data-driven investment. Observed Statistically Significant Positive Abnormal Returns (a jump in stock price) immediately after a firm announced a data-driven investment (Lee et al., 2017). Because the market is efficient at pricing these dynamic capabilities, the firm's valuation reflects its anticipated ability to adapt and grow in data-rich environments, which the authors identify as a primary driver of market value beyond traditional accounting metrics. Recent scholarship continues to underscore data's role as a valuable intangible asset. Mohan et al. provide a comprehensive review of enterprise data valuation methods and conclude that traditional accounting frameworks systematically undervalue data-driven capabilities in firm valuation (Mohan et al., 2026). Similarly, Kvalvik highlights the unique economic properties of data (non-rivalry, scalability) and the resulting challenges for market pricing, particularly in smaller economies where network effects are limited (Kvalvik et al., 2026).

Under DC logics valuation becomes more forward-looking, more expectation-driven, and less tightly coupled to contemporaneous financial indicators. Firms positioned higher along the data centrality spectrum may therefore attract higher valuation multiples independent of their current revenue base, reflecting investor beliefs about future extractive and combinatorial potential rather than realized output. Yet most existing studies focus on large, technology-dense economies, leaving open the question of whether similar valuation dynamics emerge in smaller or peripheral markets, where informational frictions and institutional constraints differ markedly.

2 METHODOLOGY

The research design employs a hybrid analytical approach, triangulating non-parametric trend assessments with parametric econometric estimations to ensure the robustness of the findings across varying statistical assumptions.

Data and Empirical Strategy

To test the hypothesis that higher-tier Data Capitalists command a valuation premium independent of revenue, the empirical dataset was engineered from the "TOP 101 of Latvia's Most Valuable Enterprises" list for the year 2024. Published by the leading Latvian corporate finance company "Prudentia" in collaboration with the official stock exchange, Nasdaq Riga (top101.lv), this ranking provides a comprehensive cross-section of the most economically significant firms operating in Latvia. The sample spans all key economic sectors—such as IT & Telecommunications, Manufacturing, Consumer Trade, and Financial Services—encompassing 17 distinct industries. For the quantitative analysis, key financial indicators—including turnover, firm value (in Million EUR), EBITDA, profitability ratios, and year-over-year growth—were extracted to construct the dependent and independent variables necessary for the valuation models.

All analyses were performed using the Python programming language (version 3.13.0). The study relied on several core libraries, including *pandas* and *numpy* for data manipulation and calculation, *statsmodels* and *linearmodels* for econometric modelling (specifically PooledOLS and Random Effects), and *scipy* for non-parametric statistical testing.

Python was selected against alternatives such as R or Stata primarily for its superior flexibility in handling the data engineering phase of this research project, and its robust ecosystem for panel data analysis. The ability to script a reproducible workflow from raw data ingestion to advanced regression within a single environment was a decisive factor in ensuring the transparency and replicability of the empirical process.

Operationalisation of the Data Capitalism Continuum

To apply theoretical insights empirically, it is insufficient to merely distinguish between 'digital' and 'traditional' firms. Such a binary approach fails to capture the nuance of how data is utilized across different business models. Consequently, this study deploys a novel tool—the Data Capitalism Maturity Framework (DCMF)—to categorise firms into four tiers based on four key dimensions with different weights: data extraction (20%), data commodification (20%), data exploitation (30%), and data centrality (30%).

The classification logic proceeds as follows:

- Primary Data Capitalists (Tier 3): The apex of the scale. For these firms, data extraction, commodification, and exploitation are central to the business model, with data often serving as the primary product or a core revenue stream.
- Secondary Data Capitalists (Tier 2): These firms significantly leverage data for competitive advantage and possess advanced analytics capabilities, yet data is not their primary product or value driver.
- Tertiary Data Capitalists (Tier 1): Firms that utilize data primarily to bolster operational efficiency, engaging in basic data collection and limited data-driven decision-making within traditional business models.
- Non-Data Capitalists (Tier 0): Firms showing minimal evidence of leveraging data as a significant value driver, relying instead on traditional value creation mechanisms with limited digital transformation.

The classification process involved a mixed-methods approach integrating qualitative analysis with a "Human-in-the-Loop" (HITL) methodology. This AI-assisted methodology was implemented following approaches described by Al-Turki et al. (2024) regarding RASE tagging in compliance building and Umakanth (2025) regarding validating GenAI outputs in clinical settings. Human-in-the-Loop systems position human agents as integral components within the decision-making pipeline, where human judgment directly influences the system's operational flow. The human element serves as an active participant in the computational process, providing real-time feedback, validation, or decision-making capabilities.

The iteration of DCMF logic proceeded in two blocks. First, the author conducted a systematic review of public documents—corporate websites, annual reports (2021–2023), and press releases for each firm. These materials were read, annotated, and assembled into "intelligence folders". Firms were manually assessed against the four dimensions described earlier. Each dimension was scored on a four-point ordinal scale (0 = Minimal/None to 3 = Advanced/Extensive). The scoring protocol was conservative, assigning zero in the absence of clear evidence. The composite DCMF Index Score was calculated as follows:

$$\text{DCMF\_Index\_Raw} = (Score_{Extraction} \text{ x } 0.20) + (Score_{Commodification} \text{ x } 0.20) + (Score_{Exploitation} \text{ x } 0.30) + (Score_{Centrality} \text{ x } 0.30) \quad (1)$$

The composite score is then normalised to a percentage scale:

$$DCMF\_Index\_Score = \text{Round}\left(\frac{\text{DCMF\_Index\_Raw}}{3.0} \times 100\right) \quad (2)$$

This formulation ensures transparency and replicability of the classification procedure.

Firms were then categorised into one of four mutually exclusive tiers based on a sequential logic:

- Tier 0 (Non-Data Capitalist): Data Extraction score is 0 (Gatekeeper Rule).
- Tier 3 (Primary Data Capitalist): (Score ≥ 2 in both Extraction and Commodification) OR (Score = 3 in both Extraction and Exploitation).
- Tier 2 (Secondary Data Capitalist): Remaining entities with DCMF Score ≥ 51.
- Tier 1 (Tertiary Data Capitalist): Remaining entities with DCMF Score ≤ 50.

To ensure reliability and mitigate bias, the classification was validated using two advanced LLMs: Google's Gemini 2.5 Pro and OpenAI's ChatGPT-O3. Both LLM models were provided with the data from "intelligence folders" and a standardised prompt with instructions and formula to perform an independent classification. In cases of discrepancy, a reconciliation process was triggered involving a re-examination of the evidence. The final classification decision in all cases rested with the human researcher, ensuring that AI remains only a tool for calibration rather than an autonomous decision-maker.

Analytical Approach

The empirical analysis is structured in two distinct stages to validate the research hypothesis from both a static and longitudinal perspective. The analysis was conducted in two distinct stages to ensure robustness. The first stage utilizes the 2024 cross-sectional dataset to establish baseline relationships. Given that the DC_Tier is an ordinal variable and financial valuation multiples often exhibit non-normal distributions and heteroscedasticity, we first employ non-parametric statistical testing. Specifically, Spearman's rank correlation (ρ) is used to detect monotonic trends between a firm's DCMF tier and its valuation multiple. This distribution-free approach ensures that any observed relationship (or lack thereof) is not an artifact of outliers or specific functional form assumptions.

The second stage expands the scope to a panel dataset, allowing for a more sophisticated econometric treatment of the data. Ranking Year is used as the time dimension for panel analysis. This aligns the market valuation (dependent variable) with the specific historical and macroeconomic context in which investors assessed the firms. Since market valuation is a forward-looking metric that consumes the previous year's financial activity (Activity Year) as an input, the Ranking Year represents the most accurate 'moment of judgment' for testing market logic.

To isolate the "Data Effect" from other firm characteristics, log-log regression model was estimated via Pooled Ordinary Least Squares (OLS) and Random Effects (RE) models. Industry effects are absorbed as time-invariant heterogeneity within the Random Effects structure. The model is specified as follows:

$$Log(Valuation\_Multiple)_{it} = \beta_0 + \beta_1(DC\_Tier)_{it} + \beta_2(Log\_Turnover)_{it} + \gamma(Industry)_i + \alpha_i + \epsilon_{it} \quad (3)$$

The variables in the equation map to the analysis as follows:

- $Log(Valuation\_Multiple)_{it}$: Dependent variable (log_multiple).
- β_0: The const (Intercept)
- $\beta_1(DC\_Tier)_{it}$: The "Data" effect.
- $\beta_2(Log\_Turnover)_{it}$: The "Size" effect.
- $\gamma(Industry)_i$: Industry dummies. It proves that the lack of a Data Premium isn't just because of the sectors the firms are in—it exists across sectors.
- α_i: The Individual (Firm) Effects. Random Effects (RE) estimator in the code, was accounted for the fact that each company has its own unique "baseline" value that doesn't change much over time.
- ϵ_{it}: "Idiosyncratic Error," or the noise in the data for each specific year.

This two-stage hybrid design allows the analysis to isolate whether a systematic valuation premium emerges along the Data Capitalism continuum, or whether observed market value is primarily a function of firm size and sectoral composition within the specific institutional and market constraints of a small, open economy like Latvia.

3 RESULTS AND DISCUSSION

What follows is not a hunt for confirmation. If anything, it is a slow dismantling of an expectation. The intuition motivating this paper—that firms which treat data as something close to capital might be rewarded accordingly by the market—feels reasonable, almost obvious. Yet once the numbers are laid out, that intuition begins to wobble. The analysis moves back and forth between simple inspection and more formal testing, not because the former is naïve and the latter sophisticated, but because each reveals different fractures.

3.1 Descriptive Patterns and Non-Parametric Correlation Analysis

The application of the DCMF to the 2024 cross-sectional dataset reveals a highly uneven distribution of data maturity among Latvia's most valuable firms. The majority of firms from top 101 occupy the lower and intermediate tiers of the framework, with Tertiary Data Capitalists (Tier 1) forming the largest cohort (48 firms). Non-Data Capitalists (Tier 0) and Secondary Data Capitalists (Tier 2) follow with 19 and 23 firms, respectively, while Primary Data Capitalists (Tier 3)—those for whom data extraction, commodification, and exploitation are structurally central to the business model—represent only a small elite segment of 11 firms.

When valuation multiples are examined across these tiers, the resulting pattern departs sharply from the monotonic structure one would expect if markets systematically capitalised data maturity. As summarised in Table 1, Tier 1 firms exhibit the highest mean valuation multiple (1.989), marginally exceeding even the most data-intensive Tier 3 firms (1.920). In contrast, Tier 2 firms—despite possessing advanced data capabilities and analytics functions—display a pronounced decline in valuation, with a mean multiple of only 1.406. Median values reinforce this irregular structure: Tier 3 firms show relatively strong median multiples (1.804), while Tier 1 and Tier 2 firms cluster around substantially lower medians.

These descriptive statistics already suggest that the relationship between data maturity and market valuation is neither linear nor hierarchical. Firms that rely minimally on data as a value driver are, on average, valued as highly as—or even more highly than—those operating at the apex of the DC continuum. Rather than reflecting a smooth progression from "traditional" to "data-centric" valuation, the distribution appears fractured, hinting at the absence of a coherent market logic linking data centrality to firm value.

Tab. 1 Descriptive Statistics of Valuation Multiple by DC Tier (2024)

DC Tier	Count (N)	Mean Multiple	Median Multiple
Tier 0 (Non-DC)	19	1.958	1.407
Tier 1 (Tertiary)	48	1.989	1.032
Tier 2 (Secondary)	23	1.406	1.068
Tier 3 (Primary)	11	1.920	1.804

Source: Author's design based on Python output.

To see whether this visual disorder conceals a subtler form of order, a non-parametric test was applied. Spearman's rank correlation, chosen precisely because it is well suited to the ordinal nature of the DC_Tier variable and avoids imposing functional form assumptions on the valuation data. The results confirm the descriptive intuition. The correlation between DC_Tier and valuation multiple is weak, negative, and statistically insignificant ($\rho = -0.1054$, $p = 0.2943$). In practical terms, this means there is no evidence of a monotonic increase in valuation multiples as firms move up the Data Capitalism tiers. The null hypothesis—that there is no association—cannot be rejected. The market does not appear to systematically reward higher data maturity with higher multiples, at least not in a way that survives basic statistical testing.

While Spearman's correlation tests the ordinal trend, it does not account for the confounding effects of firm size or industry. To isolate the specific impact of the DC_Tier on valuation, the analysis proceeded to a multivariate regression framework using the panel dataset (2022–2024). The Random Effects model was selected to account for unobserved heterogeneity—specifically, the unique, time-invariant characteristics of each firm that might influence its valuation baseline. The model controlled for firm size (log of turnover) and industry fixed effects, effectively asking: if two firms in the same industry had the same revenue, but different data tiers, would the market value them differently?

3.2 Multivariate and Panel Data Estimation

Descriptive patterns can mislead, or at least exaggerate. A few unusually valued firms can tilt an average; medians calm things down but don't explain much. To move past surface impressions, the analysis turns to multivariate estimation—not in the hope of "revealing" a hidden effect, but to see whether one survives once the obvious confounders are taken seriously. The first pass uses the 2024 cross-section alone. Valuation multiples are log-transformed, turnover enters in logs as a crude proxy for scale, and industry differences are absorbed through categorical controls. The question is almost embarrassingly simple: if two firms operate in the same sector and generate comparable revenue, does being further along the Data Capitalism continuum buy any extra multiple at all?

The answer, at least in 2024-year snapshot, is no. The coefficient on DC_Tier is effectively zero—numerically small, statistically empty. It does not wobble near conventional significance thresholds. Table 2 makes this visible at a glance. In the 2024 OLS specification, the estimated effect of moving up one data tier is effectively zero (0.0014, p=0.9859). If a data premium exists, it is not merely modest; it is undetectable.

What does emerge, and does so forcefully, is the role of scale. Turnover enters with a negative and highly significant coefficient. Larger firms, all else equal, trade at lower multiples. This is not a one-off curiosity of the cross-section but a pattern strong enough to dominate the model. One could read this as a preference for compactness, for focus, perhaps even for firms that look legible and containable in a small market. Or, more cynically, as a discount applied to business models that chase volume rather than margin. Either way, it already complicates the popular story in which data-driven firms are rewarded precisely because they scale.

Tab. 2 Comparative Regression Results (2024 OLS vs. 2022-2024 Panel RE)

Variable	2024 Cross-Section (OLS)	2022–2024 Panel (Random Effects)
DC_Tier (Effect)	0.0014 (p=0.9859)	0.0757 (p=0.2322)
Log_Turnover (Size)	-0.4395* (p=0.0000)	-0.4595* (p=0.0000)
Industry Controls	Included	Included
R-Squared (Overall)	0.8261	0.6345

Source: Author's deign based on python output. Note: *** denotes significance at the p < 0.01 level.

To assess whether this finding reflects a short-term anomaly or a more persistent structural pattern, the analysis is extended to a panel dataset covering the period 2022–2024. Both Pooled OLS and Random Effects estimators are employed, with the Random Effects model serving as the preferred specification due to its ability to account for unobserved, time-invariant firm heterogeneity while retaining cross-sectional variation. The panel results closely mirror the cross-sectional findings. Across all specifications, DC_Tier remains statistically insignificant, with coefficients that are small in magnitude and unstable across models. In the panel RE model reported in Table 2, the point estimate does rise somewhat (to 0.0757), enough to suggest a direction but not enough to command belief. The p-value (p=0.2322) stays comfortably above any conventional threshold. The market, viewed over multiple years, still does not appear to price data maturity as an independent attribute.

By contrast, the size effect not only persists but sharpens. The negative association between turnover and valuation multiples survives the transition to panel data almost unchanged. The dominant driver of valuation is Log_Turnover. The coefficient of -0.4595 in the panel model reveals a persistent "Revenue Penalty": for every 1% increase in turnover, a firm's valuation multiple decreases by approximately 0.46%. This suggests that the local market values "niche" scale and tangible asset bases over the high-turnover, scalable models characteristic of DC.

3.3 Discussion: Market Blindness to Data Maturity

Higher data-maturity tiers are not associated with higher valuation multiples once size and industry are controlled for. This null finding is not a statistical artefact: it emerges consistently across non-parametric tests, cross-sectional OLS, and panel Random Effects models. Rather than indicating market inefficiency, the absence of a premium appears to reflect the specific institutional and scale constraints of a small, open economy. In thin capital markets with limited analyst coverage and few global exit opportunities, investors appear to rely more heavily on observable, tangible signals (size and sector) than on forward-looking intangible capabilities whose payoffs are harder to appropriate locally.

Why does the market seem "blind" to data? One plausible explanation lies in the nature of information asymmetry in a small, open economy. Unlike the US or UK markets, where analysts and investors aggressively price in intangible assets and future growth potential, the findings reveal that the Latvian capital market remains anchored in traditional industrial logic. The high R-squared (0.826) indicates that valuation is highly predictable, but that predictability is driven by Size and Industry rather than data maturity. The narrative of "optionality"—the idea that data today equals future profits—may not yet be a convincing investment thesis for local investors. They may view data-driven investments as cost centres (OpEx) rather than value-generating assets.

The magnitude and stability of the turnover coefficient are difficult to ignore. With estimates clustered between −0.43 and −0.46 across specifications, the models imply that a 1% increase in revenue is associated with roughly a 0.45% decline in the valuation multiple. Scale, rather than being rewarded, is systematically discounted. Investors do not appear to price expansion as future upside; instead, growing turnover seems to signal maturity, constraint, or diminishing marginal returns. Sectoral controls reinforce this interpretation. Firms in Natural Resources and Real Estate exhibit consistently higher baseline valuation levels, suggesting that the market continues to privilege tangible asset bases and familiar cash-flow structures. In this sense, the prevailing valuation logic remains physical-asset–heavy, with data-intensive capabilities failing to translate into a comparable premium.

The most striking implication is a persistent revenue penalty. As data-oriented firms scale their operations, they encounter multiple compression rather than the multiple expansion observed among global platform firms. The negative coefficient on log turnover implies that larger firms are viewed as slow-growth entities, even when their business models rely heavily on data-driven processes. In a small domestic market with limited expansion potential, scale appears to signal saturation rather than dominance. Consequently, higher valuation multiples accrue disproportionately to smaller firms—whether data-centric or traditional—indicating that growth scarcity crowds out any pricing of data sophistication.

In large economies, Data Capitalists like Google or Amazon possess network effects that are nearly impossible to breach. In Latvia, the scale of data is naturally limited by the population size (under 2 million). A "Primary Data Capitalist" in Latvia simply cannot accumulate the volume of behavioural data that a peer in California or London can. Therefore, the market may correctly perceive the "intangible asset" of data as less defensible and less scalable in this geographic context, refusing to award a premium for capabilities that cannot be leveraged into regional or global monopoly power.

This divergence from findings in larger economies is instructive. It suggests that DC is not a single, uniform regime whose financial logic travels intact across borders. Its translation into market value appears contingent: on market depth, on investor culture, on the size of the economic playground itself. The mechanisms of data extraction and commodification may be present, but their financial signalling remains muted. Still, none of this implies that data does not matter. It may shape margins, resilience, or strategic optionality in ways that valuation multiples fail to capture. But as far as market pricing is concerned—at least in the Latvian case—the evidence points in a consistent direction. Once size and sector are accounted for, data maturity fades into the background. What remains is a market that seems to favour contained revenue, familiar assets, and business models that fit its own scale, rather than those optimized for a different one.

CONCLUSION

This study set out to test whether the logic of Data Capitalism has made its way into market valuation in Latvia—and found little evidence that it has. Across both the 2024 cross-section and the multi-year panel, firms that rank higher on data maturity are not rewarded with higher valuation multiples; if anything, they are priced as if data were largely incidental. What the market does price, consistently and forcefully, are traditional industrial anchors—natural resources, real estate, and contained scale. The persistent negative relationship between turnover and valuation suggests a structural ceiling for growth-oriented, data-intensive firms: as they expand, their multiples compress, limiting their ability to finance further growth through valuation alone. This creates a quiet but consequential divergence. Global tech leaders rely on elevated valuations to fund expansion, experimentation, and loss-making scale; Latvian firms operating under a market logic that discounts data and penalises size are unlikely to follow the same trajectory. Several limitations temper these conclusions. The Data Capitalism Maturity Framework, while systematic, necessarily simplifies heterogeneous data practices; valuation data remain thin in a small market; and unobserved strategic factors may escape both cross-sectional and panel controls. Still, within these constraints, the pattern is remarkably stable.

Future research could usefully extend this work in three directions. First, comparative studies across other Central and Eastern European countries would clarify whether the absence of a data premium is specific to Latvia or characteristic of small open economies in the region. Second, qualitative investor interviews or analyst reports could reveal the precise heuristics used to price intangible assets in thin markets. Third, alternative performance metrics (e.g., Tobin's Q, EV/EBITDA, or long-term stock returns where available) could test whether data maturity affects value creation even if it does not affect valuation multiples. Such extensions would deepen our understanding of how Data Capitalism travels — or fails to travel — beyond the world's largest financial centres.

ACKNOWLEDGEMENT

During the preparation of this work, the author used ChatGPT, an AI language model developed by OpenAI, to improve the readability and grammatical accuracy of the text. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

REFERENCES

Al-Turki, D., Hettiarachchi, H., Gaber, M. M., Abdelsamea, M. M., Basurra, S., Iranmanesh, S., Saadany, H., & Vakaj, E. (2024). Human-in-the-loop learning with LLMs for efficient RASE tagging in building compliance regulations. *IEEE Access, 12*, 185291–185306. https://ieeexplore.ieee.org/document/10778545

Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society, 8*(1), https://doi.org/10.1177/20539517211017308

Bolin, G. (2022). The value dynamics of data capitalism: Cultural production and consumption in a datafied world. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *New perspectives in critical data studies: The ambivalences of data power* (pp. 167–186). Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0_8

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). *Strength in numbers: How does data-driven decision-making affect firm performance?* SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1819486

Cheong, H., Kim, B., & Vaquero, I. U. (2023). A data valuation model to estimate the investment value of platform companies: Based on discounted cash flow. *Journal of Risk and Financial Management, 16*(6). https://doi.org/10.3390/jrfm16060293

Damodaran, A. (2009). *Valuing young, start-up and growth companies: Estimation issues and valuation challenges* (SSRN Scholarly Paper No. 1418687). Social Science Research Network. https://doi.org/10.2139/ssrn.1418687

Damodaran, A. (2025). *Investment valuation: Tools and techniques for determining the value of any asset* (4th ed.). Wiley.

Ker, D., & Mazzini, E. (2020). *Perspectives on the value of data and data flows* (OECD Digital Economy Papers No. 299). OECD Publishing. https://ideas.repec.org/p/oec/stiaab/299-en.html

Kvalvik, P., Misra, S., Mishra, B., & Dash, S. (2026). Data as an asset: Economic traits, market dynamics, and monetization challenges. *WIREs Data Mining and Knowledge Discovery, 16*(1), e70072. https://doi.org/10.1002/widm.70072

Lee, H., Kweon, E., Kim, M., & Chai, S. (2017). Does implementation of big data analytics improve firms' market value? Investors' reaction in stock market. *Sustainability, 9*(6), 978. https://doi.org/10.3390/su9060978

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*. https://hbr.org/2012/10/big-data-the-management-revolution

Mohan, S. K., Bharathy, G., & Jalan, A. (2026). Enterprise data valuation—A targeted literature review. *Journal of Economic Surveys, 40*(1), 73–92. https://doi.org/10.1111/joes.12705

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. PublicAffairs. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media-1.carnegiecouncil.org/import/studio/To_Save_Everything_Click_Here.pdf

Nissim, D., & Penman, S. H. (2001). Ratio analysis and equity valuation: From research to practice. *Review of Accounting Studies, 6*(1), 109–154. https://doi.org/10.1023/A:1011338221623

Olabode, O. E., Boso, N., Hultman, M., & Leonidou, C. N. (2022). Big data analytics capability and market performance: The roles of disruptive business models and competitive intensity. *Journal of Business Research, 139*, 1218–1230. https://doi.org/10.1016/j.jbusres.2021.10.042

Pfeiffer, S. (2022). *Digital capitalism and distributive forces*. transcript Verlag. https://www.scienceopen.com/hosted-document?doi=10.13169/prometheus.39.2.0130

Srnicek, N. (2017). The challenges of platform capitalism: Understanding the logic of a new business model. *Juncture, 23*(4), 254–257. https://doi.org/10.1111/newe.12023

Tambe, P. (2014). Big data investment, skills, and firm value. *Management Science, 60*(6), 1452–1469. https://doi.org/10.1287/mnsc.2014.1899

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.