Data Rent and Surplus Value in the Digital Economy*

Siavash Mohades[1,2] Maria Savona[1,3]

[1]LUISS Guido Carli, [2]Maastricht University, [3]University of Sussex,

January 19, 2026

Abstract

The commonly referred-to concept of "data as the new oil" quickly dismissed a a central theoretical question: What is the nature of data as production factor; how data generates value, and who captures it and through what mechanisms. We develop a formal model demonstrating that data operates through two analytically distinct channels with opposing distributional implications. First, data enhances productivity through improved prediction and coordination, genuine value creation that would persist under any ownership arrangement. Second, data enables rent extraction: exclusive control allows platforms to capture value they did not create, extracting payments from data, generating users and data-dependent firms. We formalise this distinction by adapting Marx's categories of ground rent to informational assets, distinguishing differential data rent (arising from variations in data quality and quantity) from absolute data rent (arising from monopolistic control of non-reproducible assets). The model yields three main results. First, data sharing across firms preserves productivity gains while eliminating monopoly rents, establishing that data's productive value does not require private ownership. Second, privacy protection and data sharing function as policy substitutes: both reduce platform rents, but only data sharing preserves productivity. Third, calibrating the model to GDPR evidence explains why privacy regulation increased market concentration by 17% while reducing activity by 12%: it raised data scarcity without democratising access. The true policy trade-off is not between privacy and productivity but between privacy and rent extraction. Strong data, sharing requirements combined with privacy protection could maintain productivity while eliminating rents; current policy achieves neither.

Keywords: Data economy, rent extraction, platform capitalism, Marxian economics, surplus value, privacy regulation, data sharing, digital monopoly

JEL Codes: L12, L86, O34, P16, B51

1 Introduction

The proposition that "data is the new oil" has become a ubiquitous framing for understanding value creation in contemporary capitalism. Yet this analogy, while evocative, fundamentally mischaracterises data's economic nature. Oil is rival and exhaustible; data is nonrival and inexhaustible. Oil must be discovered; data is produced continuously as a byproduct of human activity. Most critically, the oil metaphor obscures a central political-economic question: not merely how data generates value, but who captures the value data generates and through what mechanisms. This paper develops a formal model demonstrating that data functions less as a productive input receiving its marginal contribution than as a mechanism for rent extraction, a reconceptualisation with profound implications for understanding digital capitalism's distributional dynamics and for designing effective regulatory interventions.

*This document is an early-stage draft. Please do not cite. Contact s.mohadesforooshani@maastrichtuniversity.nl for the latest version.

The economic significance of data has grown enormously over the past two decades. Digital platforms now constitute seven of the world's ten most valuable companies by market capitalisation, and their valuations derive substantially from proprietary data assets and the algorithmic capabilities these enable. Corrado et al. (2022) have documented how intangible capital, encompassing software, data, organisational processes, and intellectual property, now exceeds tangible capital in most advanced economies, fundamentally reshaping growth accounting and productivity measurement. Jones and Tonetti (2020) established that data's nonrivalry creates increasing returns and efficiency gains when data flows freely, yet observed that firms systematically hoard data rather than share it. This behavioural puzzle, rational for individual firms yet collectively suboptimal, points toward market structures that reward data exclusivity over data productivity, a pattern consistent with rent-seeking rather than competitive factor markets.

Mainstream economics has approached data primarily through the lens of production theory, treating data as another factor entering production functions. In Farboodi and Veldkamp (2021)'s model of data economy, data accumulates like capital in a Solow growth model, depreciates at measurable rates, and enhances firm productivity through improved prediction. These predictions of such theoretical models were also tested empirically across various sectors of the economy (Mohades and Savona, 2025; Beraja, Yang and Yuchtman, 2023), where they quantify the effect of data access on firms' outcomes, treating data as a measurable input with identifiable effects. These contributions advance our understanding of the productive role of data, yet their theoretical frameworks necessarily obscure distributional questions. When data is modelled symmetrically as a production factor, the platform's data-derived income appears as a legitimate factor payment rather than rent extraction.

This analytical gap matters because mounting empirical evidence suggests data markets exhibit precisely the characteristics that classical economists associated with rent-generating assets: monopolistic control, barriers to entry, and returns disconnected from productive contribution. De Loecker, Eeckhout and Unger (2020) document that aggregate markups rose from 21% above marginal cost in 1980 to 61% by 2016, with increases concentrated among firms at the distribution's upper tail. Autor et al. (2020) show that industry concentration has increased substantially across U.S. sectors, with "superstar firms" capturing growing market shares while employing proportionally less labour, mainly attributed to digital technologies' scale economies. Philippon (2019) demonstrates in his comprehensive study that American markets have become systematically less competitive since 2000, with European markets, long derided for regulatory sclerosis. These findings describe an economy where factor shares increasingly favour owners of scarce, excludable assets rather than contributors to productive output.

The theoretical challenge is to explain why data generates rents rather than merely competitive returns, and how these rents relate to classical categories of exploitation and extraction. Acemoglu et al. (2022) provide one mechanism: data externalities. When individuals share data with platforms, they reveal information about other individuals, depressing data prices below competitive levels and generating excessive data sharing that benefits platforms at users' collective expense. Posner and Weyl (2018)'s "data as labor" framework identifies another: users provide valuable productive inputs: attention, behavioral information and content, all without compensation, a relationship they characterize as "techno-feudalism." Acquisti, Taylor and Wagman (2016)'s survey of privacy economics documents systematic information asymmetries that prevent consumers from making welfare-maximizing data decisions. These contributions illuminate specific market failures but lack a unified framework that connects data's roles in production, distribution, and accumulation.

This paper provides that framework by reconceptualising data through classical Marxian political economy. Our core theoretical contribution is to demonstrate that data operate through two analytically distinct channels, with opposing distributional implications. First, data enhances productivity: better information enables more efficient matching, reduced uncertainty, and improved prediction across economic activities. This productivity channel represents genuine value creation and would persist under any ownership arrangement. Second, data enables rent extraction: exclusive control over data assets allows platforms to capture value they did not create, extracting payments from both data-generating users and data-dependent businesses. This rent channel represents distributional transfer rather than value creation and depends critically on ownership structures that permit monopolistic exclusion.

We formalise this distinction by adapting Marx's categories of ground rent to informational assets. Just as landlords extract differential rent from variations in soil fertility and location, platforms extract differential data rent from variations in data quality, comprehensiveness, and algorithmic sophistication. Superior data assets enable platforms to offer better services, attracting more users who generate more data in a self-reinforcing cycle that Srnicek (2017) terms the "network effect moat." Just as landlords extract absolute rent from monopoly ownership of a non-produced input necessary for production, platforms extract absolute data rent from control over data assets that competitors cannot replicate, regardless of their productive efficiency. Christophers (2020)'s account of contemporary "rentier capitalism" documents how such rent extraction has become central to corporate strategy across sectors, from pharmaceuticals to finance to digital platforms.

Our model distinguishes labour exploitation from rent extraction as separate mechanisms of surplus appropriation. Following Marx's framework, we define the rate of surplus value σ as the ratio of unpaid to paid labour time. Data-enabled monitoring, algorithmic management, and just-in-time scheduling intensify exploitation by extending effective working time, reducing porosity, and transferring coordination costs to workers (Huws, 2014; Fuchs, 2014). This exploitation occurs within the employment relationship and affects the distribution between wages and profits. Rent extraction operates differently: it captures value from parties outside the platform's direct employment, including users who provide data without compensation and businesses that are required to pay for platform access. Zuboff (2019)'s concept of "behavioural surplus", which is data collected beyond service requirements and monetised through advertising, exemplifies this rent extraction from users, while platform fees charged to merchants and app developers exemplify rent extraction from businesses.

Our model generates several results with important policy implications. First, we demonstrate formally that data sharing across firms preserves productivity gains while eliminating monopoly rents. When data is treated as a common infrastructure rather than a proprietary asset, competitive markets for data-enabled services can emerge. The productivity benefits documented by Jones and Tonetti (2020) would remain; the rent extraction documented by the monopoly power literature would disappear. This establishes a sharp theoretical distinction: data's productive value does not require private ownership, though private owners naturally prefer institutional arrangements that conflate the two.

Second, we integrate our framework with privacy economics to show that privacy and data sharing function as policy substitutes, not complements. Privacy regulation restricts data collection, reducing both productivity and rent extraction. Data sharing requirements (data commons, interoperability mandates) maintain collection while eliminating rent through competitive access. Both policies reduce platform rents, but only data sharing preserves productivity gains. This substitutability explains an apparent paradox: privacy advocates and data monopolists often find themselves aligned against data sharing proposals, despite their opposing interests in platform rents.

Third, we calibrate the model using empirical estimates from GDPR studies to explain why privacy regulation failed to achieve its stated objectives. Goldberg, Johnson and Shriver (2024) found that GDPR enforcement reduced website pageviews and e-commerce revenue by approximately 12%, while Johnson, Shriver and Goldberg (2023) documented a 17% increase in market concentration favouring large incumbents. Aridor, Che and Salz (2023)'s analysis revealed that privacy-conscious consumers who opted out imposed externalities on remaining users, whose data became more valuable precisely because fewer alternatives existed. The pattern is consistent with our model's predictions: privacy regulation increased data scarcity without democratising data access, thereby raising barriers to entry and reinforcing incumbents' positions.[1]

Fourth, the model clarifies the true policy trade-off. The conflict is not between privacy and productivity, as commonly framed, but between privacy and rent extraction. Strong data-sharing requirements could maintain or exceed current productivity levels—since nonrival data can serve multiple firms simultaneously, while eliminating monopoly rents through competitive access. Privacy regulation, by contrast, reduces both rents and productivity by restricting data flows altogether. Framing the choice as "privacy versus innovation"

[1] Google and Facebook's market shares increased following GDPR implementation, a striking validation of the framework's analytical power.

obscures these alternatives, a framing whose persistence may reflect the political preferences of rent-collecting incumbents.

This analysis contributes to several intersecting literatures. Within heterodox economics, we provide formal mathematical and testable treatment of arguments developed qualitatively by Srnicek (2017)'s Platform Capitalism, Zuboff (2019)'s account of surveillance capitalism, and Sadowski (2019)'s analysis of data as capital. Within the rent theory literature extending from Ricardo through Marx to Harvey (2006)'s contemporary applications, we show that classical categories remain analytically powerful for understanding assets that earlier theorists could not have anticipated. Within the emerging literature on the economics of data, such as Bergemann and Bonatti (2019), we offer a distributional framework that complements their production-focused analyses.

The stakes of this theoretical reframing extend beyond academic debate. If data generate rents rather than competitive returns, then standard efficiency arguments for current ownership arrangements fail. The "marginal product" that neoclassical theory attributes to data owners reflects not their productive contribution but their monopoly position. Antitrust remedies focused on consumer prices or merger blocking address symptoms rather than causes: as long as data remains privately excludable, new monopolists will replace old ones through data accumulation dynamics. The analysis suggests that data commons and mandatory sharing represent the appropriate policy instruments, an intervention in property rights rather than market regulation, targeting rent extraction at its institutional source.

2 A Marxian Model of Data as Surplus Value and Rent

Having established the classical foundations and reviewed modern literature, we now develop a formal model of data-driven rent within the Marxian framework. Following Samuelson's rigorous treatment of Marxian economics (Samuelson, 1971, 1957) and drawing on Bowles' analysis of alternative production frameworks (Bowles, 1985), we construct a model where data appears both as a factor intensifying labour exploitation and as a source of monopoly rent. This approach is more faithful to Marx's original insights than the neoclassical Cobb-Douglas formulation, as it grounds rent in the extraction and appropriation of surplus value created by labour.

2.1 Marxian Value Theory and the Role of Data

In Marxian economics, the value of any commodity is determined by the socially necessary labour time required for its production. The total value W created in production decomposes into three parts:

$$W = c + v + s, \tag{1}$$

where c is constant capital, the value of means of production (raw materials, machinery, etc.) transferred to the product, measured in labour-time equivalents, v is variable capital, the value of labour power (wages), representing the labour-time necessary to reproduce the worker, s is surplus value, the new value created by living labour beyond v, appropriated by capitalists.

The rate of surplus value or rate of exploitation is defined as:

$$\sigma = \frac{s}{v}, \tag{2}$$

which measures how much surplus labour is extracted per unit of necessary labour. Marx argued that under capitalism, $\sigma > 0$ because workers labour beyond the time needed to reproduce their own labour power.

The rate of profit is:

$$r = \frac{s}{c+v} = \frac{\sigma}{q+1}, \tag{3}$$

where $q = c/v$ is the organic composition of capital, the ratio of dead labour (embodied in means of production) to living labour employed.

2.2 Data as a Factor Intensifying Exploitation

We introduce data D into this framework by recognising that data serves multiple functions in modern production that can lead to more exploitation. Data enables real-time monitoring and algorithmic management of workers, increasing the intensity of labour and thus the amount of value produced per labour-hour (Bowles, 1985). Surveillance technologies, productivity tracking, and algorithmic task assignment all function to extract more surplus from workers. Next, data-driven automation and AI can reduce the component of labour-time devoted to reproducing labour power (by cheapening wage goods), thereby increasing s relative to v. Furthermore, better data reduces the time capital spends in circulation (inventory management, logistics optimisation), effectively increasing the annual rate of surplus value. Moreover, data-driven optimisation can reduce waste in constant capital, improving material utilisation, extending machinery life through predictive maintenance, and optimising supply chains, thereby reducing c relative to output. Lastly, data enables firms to appropriate surplus value from other sectors through monopoly pricing, price discrimination, and network effects, analogous to merchant capital's appropriation of industrial surplus.

We formalise the primary effect by making the rate of exploitation a function of the data stock:

$$\sigma(D) = \sigma_0 + \phi(D), \tag{4}$$

where $\sigma_0 > 0$ is the baseline exploitation rate without data, and $\phi(D)$ captures the additional exploitation enabled by data, with $\phi(0) = 0$, $\phi'(D) > 0$, and $\phi''(D) < 0$ (diminishing returns).

A simple functional form is:

$$\phi(D) = \beta D^{\gamma}, \quad 0 < \gamma < 1, \tag{5}$$

so that:

$$\sigma(D) = \sigma_0 + \beta D^{\gamma}. \tag{6}$$

Note: While data also affects constant capital efficiency, we focus on the labour exploitation channel as the primary mechanism.

2.3 Data Rent as a Component of Surplus Value

Surplus value s created in production must be distributed among different claimants. In Marx's schema, s is divided into, Industrial profit π_I: retained by the industrial capitalist, Interest π_F paid to financial capital, Commercial profit π_C appropriated by merchant capital, Ground rent R_L: paid to landowners. We extend this to include data rent R_D:

$$s = \pi_I + \pi_F + \pi_C + R_L + R_D. \tag{7}$$

Data rent R_D has two components, analogous to Marx's theory of ground rent:

2.3.1 Differential Data Rent

Differential data rent arises from differences in data quality, quantity, or exclusivity across firms. Following Marx's analysis of differential rent, firms with superior data can produce at below-average costs (or above-average quality) but sell at the market-determined price. The surplus profit they realize becomes rent.

Consider two firms producing the same commodity:

- Firm A (data-rich): Has data stock D_A, exploitation rate $\sigma(D_A)$, produces value $W_A = c + v + \sigma(D_A) \cdot v$

- Firm B (data-poor): Has data stock $D_B < D_A$, exploitation rate $\sigma(D_B)$, produces value $W_B = c + v + \sigma(D_B) \cdot v$

If the market price is determined by the less efficient producer (Firm B), then Firm A's differential data rent is:

$$R_D^{\text{diff}} = [\sigma(D_A) - \sigma(D_B)] \cdot v = [\phi(D_A) - \phi(D_B)] \cdot v. \tag{8}$$

In the aggregate, if we consider a marginal firm with minimal data ($D = 0$) setting the price, then any firm with $D > 0$ earns:

$$R_D^{\text{diff}} = \phi(D) \cdot v = \beta D^{\gamma} \cdot v. \tag{9}$$

2.3.2 Absolute Data Rent

Absolute data rent arises from the monopolistic control of data as a non-reproducible asset. Even firms at the technological margin (producing under the worst conditions) may pay rent if data is essential and monopolised. This parallels Marx's absolute ground rent, which exists when private property in land creates an artificial barrier preventing free capital mobility into agriculture.

In the data economy, absolute rent emerges because:

1. Data cannot be freely reproduced, as it requires historical accumulation (user interactions, transactions)
2. Network effects create natural monopolies in data collection
3. Legal and technological barriers (IP, APIs, encryption) prevent data sharing

Absolute data rent can be modeled as a baseline payment $\bar{R}_D$ that any firm must make to access the minimum viable data infrastructure:

$$R_D^{\text{abs}} = \bar{R}_D \cdot L, \tag{10}$$

where $\bar{R}_D$ is rent per unit of labour employed (reflecting the necessity of data for modern production).

Total data rent is thus:

$$R_D = R_D^{\text{abs}} + R_D^{\text{diff}} = \bar{R}_D \cdot L + \beta D^{\gamma} \cdot v. \tag{11}$$

Analytical simplification: In what follows, we focus primarily on differential data rent, as it captures the key economic mechanism: firms with data advantages appropriate surplus from those without. Absolute rent, while conceptually important, adds a constant term that does not affect the marginal analysis of accumulation dynamics and policy interventions. Where relevant, we note that total rent includes both components.

2.4 The Marxian Rate of Profit with Data

The average rate of profit in the economy is:

$$r = \frac{S - R_D}{C + V} = \frac{S}{C + V} \cdot \frac{S - R_D}{S} = r^*(1 - \theta_D), \tag{12}$$

where $r^* = S/(C + V)$ is the rate of profit without data rent, and $\theta_D = R_D/S$ is the share of surplus appropriated as data rent.

To understand the net effect of data on profitability, we must distinguish between the surplus-enhancing effect (data increases σ) and the rent-extracting effect (data rent appropriates surplus). Let $\theta_R \in [0, 1]$ denote the rent extraction rate,the fraction of data-enabled surplus that is appropriated as rent rather than retained as profit. Then:

$$R_D = \theta_R \cdot \beta D^\gamma V, \tag{13}$$

Using equation (2) with $S = \sigma(D)V$, we can write:

$$r = \frac{\sigma(D)V - R_D}{C+V} = \frac{[\sigma_0 + \beta D^\gamma]V - \theta_R \beta D^\gamma V}{C+V} = \frac{\sigma_0 + \beta D^\gamma (1-\theta_R)}{q+1}, \tag{14}$$

where $q = C/V$ is the organic composition of capital.

This reveals the dual effect of data:

1. Positive effect: Data increases $\sigma(D)$, raising surplus value
2. Negative effect: Data rent appropriates fraction θ_R of the data-enabled surplus, reducing profit available for accumulation

When $\theta_R = 0$ (no rent extraction), data unambiguously raises the profit rate. When $\theta_R = 1$ (complete rent extraction), the data-enabled surplus is entirely appropriated as rent, leaving only the baseline σ_0 affecting profitability. Intermediate values reflect the balance between productivity gains and rent extraction.

Marx famously argued that capitalism exhibits a tendency for the rate of profit to fall as the organic composition $q = C/V$ rises with mechanization. We can examine this tendency with data.

The rate of profit from equation (14) is:

$$r = \frac{\sigma_0 + \beta D^\gamma (1-\theta_R)}{q+1}. \tag{15}$$

Taking the derivative with respect to q:

$$\frac{dr}{dq} = -\frac{\sigma_0 + \beta D^\gamma (1-\theta_R)}{(q+1)^2} < 0. \tag{16}$$

The derivative is unambiguously negative, confirming Marx's tendency for the rate of profit to fall as the organic composition rises. Data affects the magnitude of this tendency but not its direction: higher data stocks D or lower rent extraction θ_R make the numerator larger, which means the absolute rate of profit is higher, but the rate of decline with respect to q is also steeper.

This formalises an important insight: data can temporarily elevate profit rates (through increased exploitation when θ_R is low), but it cannot permanently reverse the falling tendency that arises from rising q. However, if data accumulation is rapid and rent extraction is limited, tech firms can maintain high profit rates despite high capital intensity,explaining the empirical observation of sustained high profitability among data-intensive platform firms even as their capital investments grow.

2.5 Data Sharing as a Counteracting Tendency

The dual effect of data identified in equation (14), productivity enhancement through $\sigma(D)$ versus rent extraction through R_D, suggests a potential resolution through data sharing or socialisation. We formalise this by distinguishing between aggregate data availability and proprietary data control.

Let D^{total} be the total data stock in the economy, and D_i^{prop} be firm i's proprietary data. Under monopolistic control, $D^{total} = \sum_i D_i^{prop}$ with exclusive access. Under data sharing (partial or complete), define a sharing parameter $\psi \in [0,1]$ where:

- $\psi = 0$: Full monopolization (status quo)
- $\psi = 1$: Complete data commons
- $0 < \psi < 1$: Partial sharing (e.g., mandated interoperability, open data requirements)

With sharing, each firm's effective data access becomes:

$$D_i^{eff} = D_i^{prop} + \psi \sum_{j \neq i} D_j^{prop} = D_i^{prop} + \psi(D^{total} - D_i^{prop}). \tag{17}$$

The exploitation rate for firm i depends on its effective data access:

$$\sigma_i(D_i^{eff}) = \sigma_0 + \beta(D_i^{eff})^\gamma. \tag{18}$$

However, data rent now depends on differential access, not absolute levels. Differential data rent becomes:

$$R_{D,i}^{diff} = [\sigma_i(D_i^{eff}) - \bar{\sigma}] \cdot v_i, \tag{19}$$

where $\bar{\sigma}$ is the economy-wide average exploitation rate.

As $\psi \to 1$ (full sharing), $D_i^{eff} \to D^{total}$ for all firms, so $\sigma_i \to \sigma(D^{total})$ uniformly, and thus $R_{D,i}^{diff} \to 0$. Meanwhile, absolute data rent R_D^{abs} also declines as data loses its monopolistic character.

The aggregate rate of profit under data sharing becomes:

$$r(\psi) = \frac{\sigma(D^{total})}{q+1} - \frac{R_D(\psi)}{C+V}, \tag{20}$$

where $R_D(\psi)$ is total data rent, decreasing in ψ:

$$\frac{dR_D}{d\psi} < 0, \quad \text{while} \quad \frac{d\sigma(D^{total})}{d\psi} \geq 0 \quad \text{(if sharing enables better data aggregation).} \tag{21}$$

The net effect on profitability is:

$$\frac{dr}{d\psi} = -\frac{1}{C+V}\frac{dR_D}{d\psi} > 0. \tag{22}$$

Thus, data sharing partially or fully offsets the negative rent effect while preserving the positive productivity effect. If sharing is complete ($\psi = 1$), we achieve:

$$r(\psi = 1) = \frac{\sigma_0 + \beta(D^{total})^\gamma}{q+1} > r(\psi = 0), \tag{23}$$

since the rent term vanishes while exploitation remains elevated by the full data stock.

This formalises the policy insight that socialising data does not eliminate productivity gains but merely prevents their appropriation as rent, analogous to how land value taxation captures rent without reducing land's productive use. In Marxian terms, data sharing allows the working class (or industrial capital) to retain surplus value that would otherwise be extracted as monopoly rent, while maintaining the technical advances data enables.

The contradiction is that while data sharing would increase aggregate profitability and accumulation, individual data monopolists face strong incentives to prevent sharing, creating a collective action problem that only state intervention or working-class organisation can resolve.

2.6 Privacy Protection, Data Sharing, and the Collective Action Problem

The preceding analysis demonstrated how data sharing can offset rent extraction while preserving productivity gains. However, recent economics literature reveals a fundamental tension: data sharing improves aggregate outcomes but creates privacy externalities that shift surplus from individuals to platforms. This section integrates privacy economics into the Marxian framework, showing how privacy protection and data sharing jointly determine the distribution of surplus value.

The economics literature increasingly treats privacy not as an individual preference but as a public good subject to market failures. Acemoglu et al. (2022) formalise how data externalities, where one person's data sharing reveals information about correlated others, cause systematic underpricing of privacy and excessive data collection. Because low-valuation individuals leak information about high-valuation individuals, decentralised privacy decisions lead to welfare losses. Platforms capture the value of these externalities as rent. Garratt and van Oordt (2021) model privacy explicitly as a public good, demonstrating that individuals fail to internalise how their data enables price discrimination against future consumers. This dynamic rent extraction mechanism operates across time: today's data sharing reduces tomorrow's consumer surplus, but individuals making disclosure decisions do not account for these future costs imposed on others.

Empirical evidence from GDPR implementation validates these theoretical concerns while revealing policy trade-offs. Goldberg, Johnson and Shriver (2024) documents a 12% reduction in EU website activity and revenue post-GDPR using data from 1,084 firms. Johnson, Shriver and Goldberg (2023) find that while GDPR reduced data collection, it increased market concentration by 17%, with Facebook and Google gaining share as compliance costs created barriers favouring incumbents. Aridor, Che and Salz (2023) show that privacy regulation creates adverse selection: privacy-conscious consumers exit, making remaining consenting users more valuable per capita to platforms.

This evidence suggests privacy regulation imposes real costs,reducing data availability and thus productivity,while potentially entrenching existing monopoly positions. The challenge is designing policies that protect privacy without either sacrificing productivity gains or reinforcing rent extraction by incumbents.

We extend the model by introducing a privacy protection parameter $\rho \in [0, 1]$, where:

- $\rho = 0$: No privacy protection (maximal data exploitation)
- $\rho = 1$: Complete privacy (no data collection)
- $0 < \rho < 1$: Partial privacy (e.g., differential privacy, consent requirements)

Privacy protection affects both the utility of data for intensifying exploitation and the appropriability of data rent. Following Abowd and Schmutte (2019), who model privacy-accuracy trade-offs as a production function, we specify:

$$\sigma(D, \rho, \psi) = \sigma_0 + \beta D^\gamma \cdot \underbrace{(1-\rho)^\theta}_{\text{privacy cost}} \cdot \underbrace{(1+\eta\psi)}_{\text{sharing benefit}}, \tag{24}$$

where:

- $\theta > 0$ measures how privacy protection degrades data utility (higher θ means privacy more costly)
- $\eta > 0$ captures data complementarity, shared data is more productive than proprietary data
- $\psi \in [0, 1]$ is the sharing parameter from equation (17)

Functional form justification: The power function $(1-\rho)^\theta$ captures the idea that privacy protection imposes multiplicative costs on data utility, each additional unit of privacy proportionally reduces data's effectiveness. This is consistent with differential privacy literature, where privacy budgets trade off linearly (in log space) with accuracy. The parameter θ allows calibration to empirical estimates; we derive $\theta \approx 0.25$ from our GDPR simulation in Appendix A.5, suggesting moderate sensitivity.

The term $(1-\rho)^\theta$ reflects that stronger privacy protection (higher ρ) reduces the exploitation rate, consistent with evidence that privacy regulation limits algorithmic management and surveillance (Goldberg, Johnson and Shriver, 2024). The term $(1 + \eta\psi)$ captures that data sharing increases productivity through network effects and complementarities.

For data rent, privacy protection reduces monopoly power and thus rent extraction:

$$R_D(\rho, \psi) = \beta D^\gamma V \cdot (1-\rho)^\theta \cdot (1+\eta\psi) \cdot \underbrace{(1-\psi)^\lambda}_{\text{monopoly power}}, \tag{25}$$

where $\lambda > 0$ measures how data sharing erodes monopoly rents. The term $(1-\psi)^\lambda$ reflects that as $\psi \to 1$ (full sharing), differential rents vanish as demonstrated in equation (19).

The rate of profit becomes:

$$r(\rho, \psi) = \frac{\sigma(D, \rho, \psi)}{q+1} - \frac{R_D(\rho, \psi)}{C+V}. \tag{26}$$

Substituting equations (24) and (25):

$$r(\rho, \psi) = \frac{\sigma_0 + \beta D^\gamma (1-\rho)^\theta (1+\eta\psi)}{q+1} - \frac{\beta D^\gamma (1-\rho)^\theta (1+\eta\psi)(1-\psi)^\lambda}{q+1}. \tag{27}$$

Simplifying:

$$r(\rho, \psi) = \frac{\sigma_0}{q+1} + \frac{\beta D^\gamma (1-\rho)^\theta (1+\eta\psi)}{q+1} \left[1 - (1-\psi)^\lambda\right]. \tag{28}$$

The partial derivatives reveal how privacy and sharing affect profitability:

Effect of Privacy Protection (ρ):

$$\frac{\partial r}{\partial \rho} = -\frac{\theta\beta D^\gamma (1-\rho)^{\theta-1}(1+\eta\psi)}{q+1} \left[1 - (1-\psi)^\lambda\right] < 0. \tag{29}$$

Privacy protection unambiguously reduces the profit rate because it diminishes both exploitation and rent in proportion. However, the magnitude depends on ψ: when $\psi = 0$ (no sharing), $(1-\psi)^\lambda = 1$ and the term in brackets vanishes if $\lambda \to \infty$, meaning privacy is extremely costly. When $\psi \to 1$ (full sharing), $[1-(1-\psi)^\lambda] \to 1$, so privacy costs are moderated because rent was already eliminated by sharing.

Effect of Data Sharing (ψ):

$$\begin{aligned} \frac{\partial r}{\partial \psi} &= \frac{\beta D^\gamma (1-\rho)^\theta}{q+1} \left\{\eta \left[1 - (1-\psi)^\lambda\right] + (1+\eta\psi)\lambda(1-\psi)^{\lambda-1}\right\} \\ &= \frac{\beta D^\gamma (1-\rho)^\theta}{q+1} \left\{\eta + (1-\psi)^{\lambda-1} \left[\lambda(1+\eta\psi) - \eta(1-\psi)\right]\right\}. \end{aligned} \tag{30}$$

This is unambiguously positive for $\lambda > 0$, confirming that data sharing increases profitability through two channels:

1. Productivity effect: $\eta > 0$ means shared data is more productive ($\partial\sigma/\partial\psi > 0$)
2. Rent elimination effect: Higher ψ reduces monopoly rent extraction ($\partial R_D/\partial\psi < 0$)

The net effect is positive because rent elimination dominates,sharing redistributes surplus from data monopolists to productive capital (or potentially to workers) without eliminating the technical gains from data.

Cross-derivative (Privacy-Sharing Substitutability):

$$\frac{\partial^2 r}{\partial\rho\partial\psi} = -\frac{\theta\beta D^\gamma (1-\rho)^{\theta-1}}{q+1} \left\{\eta[1 - (1-\psi)^\lambda] + (1+\eta\psi)\lambda(1-\psi)^{\lambda-1}\right\} < 0. \tag{31}$$

The negative cross-derivative indicates that privacy protection and data sharing are substitutes in terms of their impact on profitability,a key finding. When data is already widely shared (ψ high), the marginal cost of privacy protection is lower because there is less rent to lose. Conversely, when privacy is strong (ρ high), the marginal benefit of data sharing is reduced because there is less usable data to share. This substitutability implies that optimal policy should balance both dimensions rather than focusing on one exclusively.

2.7 Optimal Policy Mix: Balancing Privacy and Sharing

From a social welfare perspective, the optimal policy solves:

$$\max_{\rho,\psi} \quad r(\rho, \psi) \quad \text{subject to} \quad 0 \leq \rho, \psi \leq 1. \tag{32}$$

Limitation: This optimisation uses profit rate r as the objective, which, from a Marxian perspective, reflects capitalist interests. A fuller welfare analysis would incorporate worker welfare, reducing exploitation σ benefits workers even if it reduces profits,and privacy as an intrinsic good. The analysis here identifies policies that maximise productive efficiency while minimising rent extraction, but the complete social welfare function would weight these against worker well-being and privacy protection. We return to this normative question in the policy discussion below.

From equation (28), since $\partial r/\partial\psi > 0$ always, the optimal sharing is $\psi^* = 1$ (complete data commons). Given $\psi = 1$, equation (28) becomes:

$$r(\rho, 1) = \frac{\sigma_0}{q+1} + \frac{\beta D^\gamma (1-\rho)^\theta (1+\eta)}{q+1}, \tag{33}$$

since $[1 - (1-1)^\lambda] = 1$ for any λ. Now:

$$\left.\frac{\partial r}{\partial \rho}\right|_{\psi=1} = -\frac{\theta\beta D^\gamma (1-\rho)^{\theta-1}(1+\eta)}{q+1} < 0, \tag{34}$$

implying $\rho^* = 0$ (no privacy protection) maximises profit when data is fully shared.

This stark result reveals the fundamental tension: from a pure productivity perspective, maximising data exploitation requires no privacy. However, this ignores three critical considerations:

1. Worker welfare: Privacy protection directly limits surveillance and intensification. The exploitation rate σ measures surplus extraction from labour, so reducing σ benefits workers even if it reduces profit.

2. Externalities and commons degradation: Acemoglu et al. (2022) show that individuals under-protect privacy because they don't internalize costs imposed on others. Zero privacy leads to excessive surveillance, discrimination, and manipulation,costs not captured in r.

3. Dynamic innovation effects: Beraja, Yang and Yuchtman (2023) show data availability drives innovation but raises surveillance concerns. Too little privacy may stifle participation, reducing D itself in the long run as individuals exit data-generating activities (Aridor, Che and Salz, 2023).

2.7.1 Policy Implications: socialising Data, Not Surveillance

The model reveals a path to resolve the privacy-productivity dilemma: maximise data sharing while protecting individual privacy. This suggests policies should:

1. Mandate interoperability and data portability: Increase ψ by requiring platforms to share data with competitors and with users themselves, reducing monopoly rents while preserving aggregate data utility.

2. Implement privacy-preserving technologies: Use differential privacy, federated learning, and secure computation to enable data aggregation (high effective ψ) while protecting individual records (positive ρ). Strack and Yang (2024) provide mechanisms for revealing aggregate patterns without disclosing individual attributes.

3. Collective data governance: Establish data trusts, cooperatives, or unions that negotiate on behalf of groups, internalizing externalities that individual consent frameworks miss (Garratt and van Oordt, 2021).

4. Limit surveillance capabilities: Even with shared data, restrict uses for labour intensification and price discrimination,targeting the σ component directly rather than just limiting D.

5. Antitrust enforcement: Break up data monopolies to prevent the concentration effects documented by Johnson, Shriver and Goldberg (2023), ensuring privacy regulation doesn't entrench incumbents.

The Marxian framework clarifies what neoclassical analysis obscures: the conflict is not between privacy and productivity but between privacy and rent extraction. Data sharing for productive use (research, public services, competitive markets) can coexist with strong individual privacy protection. What cannot coexist is privacy with monopolistic surveillance capitalism, where R_D is maximised by combining maximum data extraction ($\rho \to 0$) with minimum sharing ($\psi \to 0$).

2.7.2 The Political Economy of Privacy Regulation

The model also illuminates why privacy regulation often fails or backfires. Policies that increase ρ without increasing ψ reduce both exploitation and rent, but hurt smaller firms disproportionately due to fixed compliance costs. Large platforms can amortise these costs over massive user bases and actually benefit from reduced competition (Johnson, Shriver and Goldberg, 2023).

The optimal trajectory is $(\rho, \psi) \to (0.5, 1)$,strong privacy protection combined with data socialisation. Current policy, exemplified by GDPR, moves toward $(\rho, \psi) \to (0.4, 0.15)$,moderate privacy at the cost of reduced sharing and increased concentration. This explains why GDPR achieved privacy gains but strengthened exactly the surveillance capitalists it aimed to constrain.

From a Marxian perspective, this failure is predictable: privacy regulation that operates through individual consent and firm compliance,rather than through collective ownership and democratic control of data infrastructure,cannot overcome the fundamental power asymmetries in the data economy. Just as worker ownership of factories alters class relations in ways that labour regulations cannot, collective ownership of data infrastructure would alter the political economy of information in ways that individual privacy rights cannot.

The model thus points toward more radical interventions: public data infrastructure, socialised AI development, and worker control over workplace data systems,policies that combine high ψ (data as commons) with high ρ (strong individual privacy) while fundamentally altering who captures the surplus that data enables extracting from labour.

This Marxian formulation reveals several insights obscured by neoclassical models:

1. Data as intensified exploitation: Data does not merely increase productivity in a neutral way; it specifically enables greater extraction of surplus from labour through monitoring, speedup, and deskilling (Bowles, 1985).

2. Rent as appropriated surplus: Data rent is not a payment for data's "marginal product" but rather a portion of surplus value created by labour and appropriated by those who control the data – analogous to ground rent appropriated by landlords.

3. Distribution conflict: The rise of data rent creates a three-way conflict over surplus: between workers and capital (the traditional class conflict), between industrial and financial capital (the traditional intra-capitalist conflict), and now between industrial capital and data monopolies (a new intra-capitalist conflict).

4. Tendency of profit rate: Data can temporarily counteract the falling rate of profit by boosting exploitation, but it also diverts surplus into rent, potentially reducing accumulation. This creates a contradictory dynamic similar to financialization.

5. Monopoly and centralization: The Marxian framework naturally accommodates monopoly, whereas neoclassical theory treats it as a "market failure." Data's role in creating winner-take-all dynamics fits naturally here.

6. Value appropriation across sectors: Data platforms appropriate value from productive sectors (e.g., retail, transportation, communication) without creating equivalent new value – similar to merchant capital in Marx's analysis. The transformation problem framework shows this as divergence between prices and values.

3 Results

3.1 Steady-State Analysis

To obtain our steady state condition, we set the growth equations in Appendix A.2 to zero, or $\dot{C} = \dot{V} = \dot{D} = 0$. Using the derivations so far into zero-growth conditions we obtain:

$$\frac{\delta_D D^*}{\alpha_D V^*} = \sigma_0 + \beta(D^*)^\gamma. \tag{35}$$

This is a nonlinear equation in D^* for given V^*. For $\gamma < 1$, there exists a unique positive solution. We can solve explicitly by defining $\tilde{D} = D^*/V^*$ (data per unit of variable capital):

$$\frac{\delta_D}{\alpha_D}\tilde{D} = \sigma_0 + \beta(V^*)^\gamma \tilde{D}^\gamma. \tag{36}$$

For large V^* (extensive economy), the second term dominates:

$$\tilde{D}^* \approx \left(\frac{\delta_D}{\alpha_D \beta V^{*\gamma}}\right)^{\frac{1}{\gamma-1}}. \tag{37}$$

Thus:

$$D^* \approx V^* \left(\frac{\delta_D}{\alpha_D \beta V^{*\gamma}}\right)^{\frac{1}{\gamma-1}} = \left(\frac{\delta_D}{\alpha_D \beta}\right)^{\frac{1}{\gamma-1}} V^{*\frac{1}{1-\gamma}}. \tag{38}$$

Since $\gamma < 1$, we have $1/(1-\gamma) > 1$, meaning per-worker data intensity $\tilde{D} = D/V$ increases as the scale grows. This captures network effects in data accumulation: as the economy scales up labour (V increases), the data stock grows super-linearly due to complementarities and positive feedback loops. Each additional worker both contributes to and benefits from the expanding data stock. The parameter $\gamma < 1$ ensures diminishing returns to data in generating surplus (via $\sigma(D) = \sigma_0 + \beta D^\gamma$), but does not prevent data quantity from growing faster than the labour force. This is consistent with observations of tech platforms where per-user data collection increases as the user base grows, known also as the classic "data flywheel" effect, where more users generate more data per user.

4 Conclusion

The Marxian production framework provides a theoretically coherent and politically relevant analysis of data as rent. By grounding the analysis in the labour theory of value and the extraction of surplus, we avoid the neoclassical mystification of treating data as just another symmetric factor of production. Instead, data appears in its true role: as a tool for intensifying exploitation and as a source of monopoly rent appropriated from productive labour.

This formulation connects naturally to the classical rent theories we reviewed earlier, showing that data rent operates through the same mechanisms as land rent, creating income for owners not through their productive contribution but through their control of a scarce, non-reproducible asset. The difference is that data scarcity is largely artificial (created by legal and technical barriers), making its rent even more amenable to policy intervention than natural resource rents.

The dynamic equations show that high data rent shares ($\theta_D \to 1$) can choke off accumulation, leading to a Marxian crisis of realization, surplus is appropriated as rent rather than reinvested, reducing effective demand and growth. This provides a Marxian explanation for secular stagnation in the data economy, complementing the Ricardian stationary state we derived earlier.

References

Abowd, John M., and Ian M. Schmutte. 2019. "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices." American Economic Review, 109(1): 171–202.

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. 2022. "Too much data: Prices and inefficiencies in data markets." American Economic Journal: Microeconomics, 14(4): 218–256.

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The Economics of Privacy." Journal of Economic Literature, 54(2): 442–492.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz. 2023. "The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR." RAND Journal of Economics, 54(4): 695–722.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." The Quarterly journal of economics, 135(2): 645–709.

Beraja, Martin, David Y. Yang, and Noam Yuchtman. 2023. "Data-Intensive Innovation and the State: Evidence from AI Firms in China." Review of Economic Studies, 90(4): 1701–1723.

Bergemann, Dirk, and Alessandro Bonatti. 2019. "Markets for information: An introduction." Annual Review of Economics, 11(1): 85–107.

Bowles, Samuel. 1985. "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models." American Economic Review, 75(1): 16–36.

Christophers, Brett. 2020. Rentier capitalism: Who owns the economy, and who pays for it? Verso Books.

Corrado, Carol, Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi. 2022. "Intangible capital and modern economies." Journal of Economic Perspectives, 36(3): 3–28.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The rise of market power and the macroeconomic implications." The Quarterly journal of economics, 135(2): 561–644.

Farboodi, Maryam, and Laura Veldkamp. 2021. A growth model of the data economy. National Bureau of Economic Research.

Fuchs, Christian. 2014. Digital labour and karl marx. Routledge.

Garratt, Rodney, and Maarten R. C. van Oordt. 2021. "Privacy as a Public Good: A Case for Electronic Cash." Journal of Political Economy, 129(7): 2157–2180.

Goldberg, Samuel, Garrett A. Johnson, and Scott Shriver. 2024. "Regulating Privacy Online: An Economic Evaluation of the GDPR." American Economic Journal: Economic Policy, 16(1): 325–358.

Harvey, David. 2006. Spaces of global capitalism. Verso.

Huws, Ursula. 2014. Labor in the global digital economy: The cybertariat comes of age. NYU Press.

Johnson, Garrett A., Scott K. Shriver, and Samuel G. Goldberg. 2023. "Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR." Management Science, 69(10): 5695–5721.

Jones, Charles I, and Christopher Tonetti. 2020. "Nonrivalry and the Economics of Data." American Economic Review, 110(9): 2819–2858.

Kalecki, Michał. 1968. "The Marxian Equations of Reproduction and Modern Economics." Social Science Information, 7(6): 73–79. Also published in *Marx and Contemporary Scientific Thought: Symposium*, De Gruyter Mouton, 1969, pp. 313–321.

Mohades, Siavash, and Maria Savona. 2025. "Uncertainty and Investments in Data and R&D."

Philippon, Thomas. 2019. The great reversal: How America gave up on free markets. Harvard University Press.

Posner, Eric, and Eric Weyl. 2018. Radical markets: Uprooting capitalism and democracy for a just society. Princeton University Press.

Sadowski, Jathan. 2019. "When data is capital: Datafication, accumulation, and extraction." Big data & society, 6(1): 2053951718820549.

Samuelson, Paul A. 1957. "Wages and Interest: A Modern Dissection of Marxian Economic Models." American Economic Review, 47(6): 884–912.

Samuelson, Paul A. 1971. "Understanding the Marxian Notion of Exploitation: A Summary of the So-Called Transformation Problem Between Marxian Values and Competitive Prices." Journal of Economic Literature, 9(2): 399–431.

Srnicek, Nick. 2017. Platform capitalism. John Wiley & Sons.

Strack, Philipp, and Kai Hao Yang. 2024. "Privacy-Preserving Signals." Econometrica, 92(6): 1907–1938.

Zuboff, Shoshana. 2019. "Surveillance capitalism and the challenge of collective action." Vol. 28, 10–29, Sage Publications Sage CA: Los Angeles, CA.

A Mathematical Derivations

A.1 The Marxian Transformation Problem Revisited

Following Samuelson's formalisation (Samuelson, 1971), we must address how labour values transform into monetary prices when data creates differential productivity and monopoly power.

In a competitive equilibrium without data, prices of production equalize the profit rate across sectors. With inputs valued at prices p_i and uniform wage w, the price system satisfies:

$$p_j = (1+r)\sum_i a_{ij}p_i + (1+r)l_j w, \tag{39}$$

where a_{ij} is the technical coefficient (input i per unit of output j), l_j is labour input per unit of output, and r is the uniform profit rate.

With data creating monopoly positions, the price system becomes:

$$p_j = (1+r_j)\sum_i a_{ij}p_i + (1+r_j)l_j w + \rho_j(D_j), \tag{40}$$

where r_j may vary across sectors (monopoly power prevents equalization), and $\rho_j(D_j)$ is the data rent per unit of output in sector j, increasing in the sector's data stock D_j.

To connect this to our rent structure, note that differential data rent per unit of output can be expressed as:

$$\rho_j(D_j) = \frac{\beta D_j^\gamma \cdot v_j}{y_j}, \tag{41}$$

where y_j is output per unit of labour in sector j and v_j is variable capital (wages) per unit of labour. Thus ρ_j increases with the sector's data stock D_j according to the power function specified in equation (9).

The deviation from competitive prices represents monopoly rent. If we denote competitive prices as p_j^* and actual prices as p_j, then:

$$p_j - p_j^* = \int_0^{D_j} \frac{\partial \rho_j(D)}{\partial D}\, dD = \frac{\beta\gamma v_j}{y_j}\int_0^{D_j} D^{\gamma-1}\, dD = \frac{\beta v_j D_j^\gamma}{y_j}, \tag{42}$$

which is the accumulated rent from data monopolization, consistent with equation (41).

Samuelson (Samuelson, 1971) showed that labour values can be transformed into prices through a linear system. With data creating monopoly positions, the transformation becomes:

Let λ_j be the labour value per unit of commodity j, and p_j the monetary price. In competitive equilibrium:

$$p_j = (1+r)\sum_i a_{ij}p_i + (1+r)l_j w = (1+r)(1+m)\lambda_j w, \tag{43}$$

where m is the markup over labour value.

With data monopolies:

$$p_j = (1+r_j)\sum_i a_{ij}p_i + (1+r_j)l_j w + \mu_j(D_j), \tag{44}$$

where $\mu_j(D_j)$ is monopoly markup per unit, rising with data stock.

The ratio of price to labour value becomes:

$$\frac{p_j}{\lambda_j w} = (1+r_j)(1+m) + \frac{\mu_j(D_j)}{\lambda_j w}. \tag{45}$$

Data-rich sectors have p_j/λ_j ratios significantly above the average, representing appropriation of value from other sectors.

A.1.1 Note on the Transformation to Prices

Our analysis remains in value terms throughout. In the transformation from labor values to monetary prices, data monopolies would create systematic divergences: data-rich sectors command prices exceeding their labor-value content, representing inter-sectoral appropriation through monopoly pricing. This price-value gap manifests the rent extraction formalized in equations (8)-(11). We omit the full price system analysis as it adds mathematical complexity without altering the distributional insights central to our argument.

A.2 Growth Equations

Let $C = c_I + c_{II}$ be total constant capital, $V = v_I + v_{II}$ be total variable capital, $D = D_I + D_{II}$ be total data stock, and $S = s_I + s_{II}$ be total surplus value. Define:

- $g_C = \Delta C/C$: growth rate of constant capital
- $g_V = \Delta V/V$: growth rate of variable capital
- $g_D = \Delta D/D$: growth rate of data stock

The laws of motion are:

$$\dot{C} = \alpha_C(S - R_D) - \delta_C C, \tag{46}$$

$$\dot{V} = \alpha_V(S - R_D) - \delta_V V, \tag{47}$$

$$\dot{D} = \alpha_D S - \delta_D D, \tag{48}$$

where:

- $\alpha_C, \alpha_V, \alpha_D$ are investment allocation parameters ($\alpha_C + \alpha_V + \alpha_D \leq 1$)
- $\delta_C, \delta_V, \delta_D$ are depreciation rates
- R_D is total data rent appropriated from surplus

Note that data accumulation receives investment from gross surplus S, but constant and variable capital accumulation receive $(S - R_D)$, reflecting that data rent is appropriated before industrial reinvestment.

A.3 Derivation of Steady State

$$\alpha_C(S^* - R_D^*) = \delta_C C^*, \tag{49}$$

$$\alpha_V(S^* - R_D^*) = \delta_V V^*, \tag{50}$$

$$\alpha_D S^* = \delta_D D^*. \tag{51}$$

From equation (51):

$$D^* = \frac{\alpha_D S^*}{\delta_D}. \tag{52}$$

Since $S^* = \sigma(D^*)V^*$ and using equation (6):

$$S^* = (\sigma_0 + \beta(D^*)^\gamma)V^*. \tag{53}$$

Substituting into equation (52):

$$D^* = \frac{\alpha_D(\sigma_0 + \beta(D^*)^\gamma)V^*}{\delta_D}. \tag{54}$$

For data rent, using equation (13):

$$R_D^* = \theta_R \beta(D^*)^\gamma V^*. \tag{55}$$

Combining equations (54) and (55):

$$D^* = \frac{\alpha_D}{\delta_D}(\sigma_0 V^* + \beta(D^*)^\gamma V^*). \tag{56}$$

Rearranging:

$$\frac{\delta_D D^*}{\alpha_D V^*} = \sigma_0 + \beta(D^*)^\gamma. \tag{57}$$

This is a nonlinear equation in D^* for given V^*. For $\gamma < 1$, there exists a unique positive solution. We can solve explicitly by defining $\tilde{D} = D^*/V^*$ (data per unit of variable capital):

$$\frac{\delta_D}{\alpha_D}\tilde{D} = \sigma_0 + \beta(V^*)^\gamma \tilde{D}^\gamma. \tag{58}$$

For large V^* (extensive economy), the second term dominates:

$$\tilde{D}^* \approx \left(\frac{\delta_D}{\alpha_D \beta V^{*\gamma}}\right)^{\frac{1}{\gamma-1}}. \tag{59}$$

Thus:

$$D^* \approx V^* \left(\frac{\delta_D}{\alpha_D \beta V^{*\gamma}}\right)^{\frac{1}{\gamma-1}} = \left(\frac{\delta_D}{\alpha_D \beta}\right)^{\frac{1}{\gamma-1}} V^{*\frac{1}{1-\gamma}}. \tag{60}$$

Since $\gamma < 1$, we have $1/(1-\gamma) > 1$, meaning per-worker data intensity $\tilde{D} = D/V$ increases as the scale grows. This captures network effects in data accumulation: as the economy scales up labour (V increases), the data stock grows super-linearly due to complementarities and positive feedback loops. Each additional worker both contributes to and benefits from the expanding data stock. The parameter $\gamma < 1$ ensures diminishing returns to data in generating surplus (via $\sigma(D) = \sigma_0 + \beta D^\gamma$), but does not prevent data quantity from growing faster than the labour force. This is consistent with observations of tech platforms where per-user data collection increases as the user base grows—the classic "data flywheel" effect where more users generate more data per user.

A.4 Dynamic Accumulation with Data Rent

We now formulate the dynamics of capital accumulation when surplus value is partially appropriated as data rent. Following the Marxian reproduction schemes (as formalised by Kalecki (1968)), we must specify how surplus is allocated between accumulation and consumption.

A.4.1 Sectoral Decomposition

Consider a two-sector economy:

Sector I: Production of means of production (constant capital goods)

$$W_I = c_I + v_I + s_I, \tag{61}$$

$$s_I = \sigma(D_I) \cdot v_I = (\sigma_0 + \beta D_I^\gamma) v_I. \tag{62}$$

Sector II: Production of consumption goods (for workers and capitalists)

$$W_{II} = c_{II} + v_{II} + s_{II}, \tag{63}$$

$$s_{II} = \sigma(D_{II}) \cdot v_{II} = (\sigma_0 + \beta D_{II}^{\gamma})v_{II}. \tag{64}$$

Total value produced in the economy:

$$W = W_I + W_{II} = (c_I + c_{II}) + (v_I + v_{II}) + (s_I + s_{II}). \tag{65}$$

A.4.2 Reproduction Conditions

For simple reproduction (no growth), the intersectoral balance requires:

$$c_{II} = v_I + s_I - R_{D,I}, \tag{66}$$

where $R_{D,I}$ is data rent extracted from Sector I. This states that the constant capital needed by Sector II equals the value available from Sector I after workers' consumption and after data rent is paid.

For expanded reproduction (accumulation), we denote:

- $\Delta c_I, \Delta c_{II}$: increments to constant capital
- $\Delta v_I, \Delta v_{II}$: increments to variable capital
- $\Delta D_I, \Delta D_{II}$: increments to data stocks
- s_I^{acc}, s_{II}^{acc}: surplus devoted to accumulation (vs. capitalist consumption)

The accumulation equations become:

$$\Delta c_I + \Delta c_{II} + \Delta D_I + \Delta D_{II} = s_I^{acc} - R_{D,I}^{acc}, \tag{67}$$

$$\Delta v_I + \Delta v_{II} = s_{II}^{acc} - R_{D,II}^{acc}, \tag{68}$$

where data accumulation now competes with physical capital accumulation for surplus value.

A.5 Calibrating the Model: GDPR as a Natural Experiment

We can use GDPR evidence to calibrate parameter values. Let the pre-GDPR state be (ρ_0, ψ_0) and post-GDPR be (ρ_1, ψ_1) with $\rho_1 > \rho_0$ (stronger privacy) and $\psi_1 < \psi_0$ (reduced data access due to market concentration effects).

Goldberg, Johnson and Shriver (2024) find a 12% reduction in activity, suggesting:

$$\frac{r(\rho_1, \psi_1)}{r(\rho_0, \psi_0)} \approx 0.88. \tag{69}$$

Assuming $\psi_0 \approx 0.2$ (limited pre-GDPR sharing), $\psi_1 \approx 0.15$ (concentration effect from Johnson, Shriver and Goldberg 2023), and $\rho_0 = 0.1$, $\rho_1 = 0.4$ (moderate increase), we can solve for θ and λ using equation (27).

Taking logarithms:

$$\ln\left(\frac{r_1}{r_0}\right) \approx \theta \ln\left(\frac{1-\rho_1}{1-\rho_0}\right) + \ln\left(\frac{1+\eta\psi_1}{1+\eta\psi_0}\right) + \ln\left(\frac{1-(1-\psi_1)^{\lambda}}{1-(1-\psi_0)^{\lambda}}\right). \tag{70}$$

For $\ln(0.88) \approx -0.128$, and assuming $\eta = 0.5$ (moderate complementarity), the first two terms give:

$$\theta \ln(0.6/0.9) \approx -0.51\theta, \tag{71}$$

$$\ln\left(\frac{1.075}{1.1}\right) \approx -0.023. \tag{72}$$

If we attribute most of the decline to privacy costs and set the concentration effect term to roughly offset at ≈ 0.02, we obtain $\theta \approx 0.25$. This implies data utility is moderately sensitive to privacy restrictions,a 10 percentage point increase in privacy protection reduces data utility by approximately 2.5

A.6 Comparison with the Neoclassical Model

Our earlier Cobb-Douglas model treated data symmetrically with capital and labour as factors of production, with each receiving its marginal product. The Marxian model reveals this as ideological mystification:

- Neoclassical: $Y = K^\alpha D^\beta L^{1-\alpha-\beta}$ with data receiving βY
- Marxian: $W = c + v + s$, with $s = \sigma(D) \cdot v$ and $R_D = \phi(D) \cdot v$

In the Marxian view:

- labour creates all new value $(v + s)$
- Capital merely transfers its value (c) to the product
- Data enables greater extraction of surplus from labour
- Data rent is appropriated from surplus, not "earned" by data's productivity

The neoclassical β obscures that data's "contribution" comes from intensifying workers' exploitation, not from data itself producing value. The Marxian $\phi(D)$ makes this explicit: data increases σ, the rate at which surplus is extracted from labour.

A.7 Policy Implications

The Marxian framework suggests different policy interventions than neoclassical analysis:

1. Data commons: Since data rent is appropriated surplus rather than earned income, socialising data (open data mandates, public data infrastructure) would not reduce "incentives" but simply prevent rent extraction – analogous to land value taxation or nationalization.
2. Worker data rights: If data increases exploitation, workers have a claim to the surplus enabled by data about them – suggesting data dividends or collective bargaining over algorithmic management.
3. Antitrust: Monopoly data rent represents value appropriation from other sectors. Breaking up data monopolies would redistribute surplus to productive capital and potentially to workers, not merely improve "efficiency."
4. Decommodification: Essential data services (search, communication, social infrastructure) could be provided publicly, eliminating rent extraction – similar to how public utilities prevent monopoly rent in electricity or water.
5. labour regulation: Limits on algorithmic management, monitoring, and intensification directly address data's role in increasing σ, which neoclassical models cannot conceptualize.