Economics Bulletin

Volume 46, Issue 1

Data processing growth model

Teemu Pekkarinen
University of Vaasa

Abstract

This paper develops a growth model in which new ideas result from people processing data. By distinguishing between ideas and data, the model provides a transparent framework for studying how information technology affects economic growth. Information technology is decomposed into three components: data processing, data generation, and data retention. The model has two candidate balanced growth path regimes of per-capita output: (i) a path in which the long-run growth rate is governed by data-processing capacity, while changes in data generation and retention affect levels but not the growth rate; (ii) a path in which long-run growth is jointly determined by data-processing and data-retention capacities. Which balanced growth path the economy obtains depends on the strength of data-processing capacity. The central insight is that improvements in data generation or retention primarily expand data stocks and have limited implications for long-run growth unless accompanied by improvements in data-processing capacity.

I am grateful to my friend and colleague Oliver Snellman for encouraging me to complete this idea and paper.
Citation: Teemu Pekkarinen, (2026) "Data processing growth model", *Economics Bulletin*, Volume 46, Issue 1, pages 334-341
Contact: Teemu Pekkarinen - pekkariselle@gmail.com.
Submitted: January 16, 2026. **Published:** March 30, 2026.

1 Introduction

Economic growth is driven by the creation of ideas. In modern economies, this process increasingly depends on the accumulation and use of data. Advances in digital technologies, artificial intelligence (AI), and data infrastructure have dramatically expanded society's capacity to generate, process, and retain data, raising the question of how these advances translate into long-run economic growth.

This paper develops a growth model that distinguishes sharply between *data* and *ideas*. Data is the effective stock of usable, codified information that can be stored and reused in idea creation, including records, observations, and corpora. It grows as economic activity generates new information, and it remains usable over time through retention and improvements in curation that slow obsolescence. Ideas, in contrast, are implementable innovations such as new technologies, production methods, or organizational blueprints that raise productivity once discovered. The economy does not invest directly in ideas; instead, it generates and retains data, which researchers process into new ideas. By separating these two objects, the model makes it possible to study how different components of information technology affect economic growth.

Information technology is decomposed into three margins: *generation, retention, and processing*. Data generation governs how much new data is produced in the economy; data retention governs how persistent and usable data remains over time (capturing both obsolescence and improvements in the effective data stock); and data processing capacity governs how effectively the existing data stock can be translated into new ideas. With these margins separated, the model delivers two candidate balanced growth path regimes. In one regime, when data processing is sufficiently strong, the long-run growth rate of per-capita output is governed by processing capacity, while changes in generation and retention primarily affect levels. In the other regime, when processing is weak, long-run growth is jointly determined by processing and retention because retention pins down the asymptotic growth rate of the data stock. The central implication is that improvements in data generation or retention have limited consequences for long-run growth unless accompanied by improvements in data-processing capacity.

The analysis builds on the idea-based growth tradition of Romer (1990), Aghion and Howitt (1992), and Grossman and Helpman (1993), and on later models that feature weak scale effects (e.g., Jones, 1995; Kortum, 1997; Segerstrom, 1998). In this literature, increasing the level of research effort raises the level of ideas and output, but does not permanently raise the long-run growth rate. The model shares this property: on the balanced growth path, trend growth is pinned down by the evolution of the data stock and data-processing capacity rather than by the level of research effort. Unlike standard semi-endogenous models in the spirit of Jones (1995), where sustained per-capita growth typically requires positive population growth, this framework can deliver positive long-run growth even when population growth is zero, provided data processing is operative and retention is sufficiently strong. The reason is that data accumulation and data processing expand the effective inputs into idea creation independently of population. Weak scale effects therefore arise endogenously from the joint dynamics of data accumulation, data-processing capacity, and idea creation. This mechanism is consistent with evidence (e.g., Bloom et al., 2020) that research productivity has declined. In this paper, that decline is interpreted as a widening gap between the expanding stock of available data and the economy's capacity to process it into implementable ideas. At the same time, the framework highlights that improvements in data-processing capacity can still amplify long-run growth.

Recent work has placed data at the center of growth and productivity analysis by emphasizing its nonrival nature, reuse, and incentive constraints (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2026; Cong et al., 2021, 2022). This literature highlights how data accumulation and sharing can generate increasing returns, but it often leaves implicit the transformation of raw data into economically useful ideas. This paper emphasizes a distinct but complementary friction: abundant data alone does not raise productivity. Productivity gains require data-processing capacity, defined as researchers' ability to interpret, combine, and translate data into implementable ideas. To isolate this channel, the model assumes that ideas are nonrival and freely usable once discovered, abstracting from intellectual property frictions. This stripped-down environment highlights the feedback loop from production to data, from data to ideas, and from ideas back to productivity, and it identifies data-processing capacity as a margin through which advances in education, automation, and artificial intelligence can durably raise economic growth.

2 Data Processing Growth

This section develops a data-processing-driven growth model and characterizes its balanced growth paths. Time $t \in [0, \infty)$ is continuous, and the time derivative of a variable X_t is denoted by $\dot{X}_t := \partial X_t / \partial t$. The growth rate of X_t is $g_X := \dot{X}_t / X_t$. Per-capita variables are written in lowercase, $x := X/L$, where X denotes the aggregate quantity and L the labor force. Throughout, all variables are assumed to have strictly positive initial values, $X_0 > 0$.

2.1 Model

Consider an economy in which labor L is allocated between researchers L_A and producers L_Y. A constant fraction $r \in (0, 1)$ of the labor force works in research, and the remaining fraction $1 - r$ works in production. The labor force grows at the constant rate $n \geq 0$.

The economy produces a single consumption good using producers and the stock of ideas:

$$Y_t = A_t^{\sigma} L_{Yt}, \qquad \sigma > 0, \tag{1}$$

where Y_t denotes output, A_t the stock of ideas, and L_{Yt} the number of producers. The parameter σ measures the effectiveness of ideas in production.

Researchers process the existing data stock to discover new ideas that can be implemented in production. Ideas therefore evolve according to

$$\dot{A}_t = D_t^{\pi} L_{At}^{\phi} - \delta A_t, \qquad \pi, \phi \geq 0, \tag{2}$$

where $D_t^{\pi} L_{At}^{\phi}$ is the idea-production technology: research labor L_{At} is combined with the data stock D_t to generate new, usable ideas. The parameter π captures the elasticity of idea creation with respect to data—higher π means data are more effectively translated into ideas—while ϕ captures the contribution of research labor to processing and extracting ideas from data. Together, (π, ϕ) summarize the economy's *data-processing* capacity. We assume that $1 > \sigma\pi$, which ensures that the balanced growth rates are well-defined and non-negative. The term δA_t captures creative destruction: at rate $\delta \geq 0$, existing ideas become obsolete.

Data evolves through two channels: *generation* and *retention*. Formally, the data accumulation process is

$$\dot{D}_t = \gamma Y_t + \rho D_t, \qquad \gamma > 0,\ \rho \in \mathbb{R}. \tag{3}$$

The parameter γ captures the economy's *economic-activity* channel of data generation: higher γ means that a larger flow of new data is produced alongside output. The parameter ρ is the net retention rate of *usable* data—with $\rho < 0$ capturing obsolescence and $\rho > 0$ capturing improvements in the effective data stock. The forces shaping retention may originate within the economy (through changes in curation, preservation, practices, or infrastructure) or from outside it (through developments beyond the economy). Accordingly, γY_t represents the *endogenous* data-generation channel, whereas ρD_t represents the *exogenous* retention channel.

In summary, the model isolates three core dimensions of information technology:

A. **Data processing** (π, ϕ) governs how effectively existing data are transformed into new ideas through the data and research-labor inputs in idea production.

B. **Data generation** (γ) captures how economic activity produces new data.

C. **Data retention** (ρ) measures the net persistence of usable data over time, with $\rho < 0$ representing obsolescence and $\rho > 0$ representing improvements in the effective data stock.

These parameters are the central objects of interest throughout the analysis.

2.2 Long-Run Growth

We focus on balanced growth paths (BGPs), defined as trajectories along which the growth rates of (Y_t, A_t, D_t) are constant and key ratios are time-invariant.

Growth in Per-Capita Output Recall that the number of researchers is $L_A = rL$ and the number of producers is $L_Y = (1-r)L$. From (1) we find that the per-capita output is

$$y_t = (1-r)A_t^{\sigma}.$$

Along any path, its growth rate is

$$g_y = \sigma g_A. \tag{4}$$

Thus, per-capita output growth is entirely mediated by idea-based technological progress in this model.

Growth in Ideas The growth of ideas, in turn, is given by

$$g_A = \frac{D_t^{\pi} L_{At}^{\phi}}{A_t} - \delta.$$

On a balanced growth path, the ratio $\frac{D_t^{\pi} L_{At}^{\phi}}{A_t}$ is constant. Hence, δ affects this constant (and thus levels) but does not influence the long-run growth rates. This implies that

$$g_A = \pi g_D + \phi n. \tag{5}$$

Equation (5) states that the growth rate of ideas is equal to a weighted sum of the growth rate of data and the growth rate of research labor. This condition is central. It captures the idea that innovation scales with both (i) how rapidly the stock of reusable data is expanding and (ii) how fast the research workforce is expanding.[1]

Growth in Data From (3), the growth rate of the data stock is

$$g_D = \gamma \frac{Y_t}{D_t} + \rho.$$

Moreover, the long-run behavior of Y_t/D_t depends on whether data growth remains tied to output through the data–ideas–output feedback loop, or whether retention dominates asymptotically. This yields two candidate balanced growth paths.

(i) Strong data-processing path We say that data processing is *strong* if Y_t/D_t is asymptotically a positive constant. In this case, the economy features a self-sustaining feedback loop: a larger data stock generates more ideas, ideas raise output, higher output generates more data, and the process repeats, yielding a balanced long-run ratio between output and data.

On the strong data-processing BGP, the term $\gamma Y_t/D_t$ remains asymptotically relevant and data must grow at the same rate as output:

$$g_D = g_Y.$$

Using $g_Y = \sigma g_A + n$ together with the balanced growth restriction from idea accumulation $g_A = \pi g_D + \phi n$, it follows that

$$g_D^s = \frac{n(1+\sigma\phi)}{1-\sigma\pi} \geq 0, \qquad \frac{Y_t}{D_t} \underset{t\to\infty}{\longrightarrow} \frac{g_D^s - \rho}{\gamma}.$$

Hence, this regime requires $g_D^s > \rho$ so that the limiting ratio Y_t/D_t is positive. In particular, whenever there is data obsolescence ($\rho < 0$), the economy necessarily falls into the strong data-processing regime. Along this balanced growth path, the growth rates of ideas and output are pinned down by g_D^s:

$$g_A^s = \pi g_D^s + \phi n, \qquad g_y^s = \sigma g_A^s.$$

[1] Recall that the number of researchers increases at the same rate n as the whole labor force L.

(ii) Weak data-processing path We say that data processing is *weak* if $\frac{Y_t}{D_t}$ is zero in the long run. In this case, expansion of the data stock does not translate into proportional gains in production, so output grows more slowly than data and the output–data ratio vanishes asymptotically.

On the weak data-processing BGP,

$$g_D^w = \rho, \qquad \frac{Y_t}{D_t} \underset{t\to\infty}{\longrightarrow} 0,$$

so asymptotically, the generation term $\gamma \frac{Y_t}{D_t}$ becomes negligible, and data growth is pinned down by retention. Along this path, idea and output growth are therefore

$$g_A^w = \pi\rho + \phi n, \qquad g_Y^w = \sigma(\pi\rho + \phi n) + n, \qquad g_y^w = \sigma(\pi\rho + \phi n).$$

Consistency with $\frac{Y_t}{D_t} \underset{t\to\infty}{\longrightarrow} 0$ requires $g_Y^w < g_D^w$, which holds precisely when $\rho > g_D^s$ (the knife-edge case $\rho = g_D^s$ leaves Y_t/D_t constant). In particular, obtaining a weak data-processing BGP requires sufficiently strong retention, which in this model means $\rho > 0$.

Results The economy's long-run balanced growth path is determined by whether ρ is below or above the benchmark growth rate g_D^s, i.e. by whether the economy falls into the regime in which data processing is strong or weak.

Theorem 1. *The model has two candidate balanced growth path regimes:*

(i) ***Strong data-processing path*** ***($\rho < g_D^s$).*** *Along the balanced growth path with strong data processing, the long-run growth rates are*

$$g_D^s = \frac{n(1+\sigma\phi)}{1-\sigma\pi}, \qquad g_A^s = \frac{n(\pi+\phi)}{1-\sigma\pi}, \qquad g_y^s = \sigma\,\frac{n(\pi+\phi)}{1-\sigma\pi}.$$

(ii) ***Weak data-processing path*** ***($\rho > g_D^s$).*** *Along the balanced growth path with weak data processing, the long-run growth rates are*

$$g_D^w = \rho, \qquad g_A^w = \pi\rho + \phi n, \qquad g_y^w = \sigma(\pi\rho + \phi n).$$

In the knife-edge case $\rho = g_D^s$, both characterizations are consistent and $\frac{Y_t}{D_t}$ remains constant.

Interpretation Information technology enters the model through three margins: *data processing* (π, ϕ), *data generation* γ, and *data retention* ρ. Processing governs how effectively the existing data stock is turned into new ideas; generation controls how much new data is produced alongside current output; and retention determines the net persistence of usable data over time ($\rho < 0$ obsolescence, $\rho > 0$ net improvements in usability). These margins affect growth through

$$g_A = \pi g_D + \phi n \quad \text{and} \quad g_y = \sigma g_A,$$

which puts data processing at the center: it governs how any change in data accumulation feeds into idea growth and, ultimately, into per-capita output growth.

When data processing is *strong*, the data–ideas–output feedback loop remains operative, and data growth is tied to output growth. In this regime, higher processing capacity π or ϕ raises trend growth by strengthening the translation of data (and research effort) into idea creation, whereas changes in γ and ρ primarily affect levels on the balanced growth path and the speed of approach toward it.

When data processing is *weak*, the generation term becomes asymptotically negligible, and data growth is pinned down by retention, $g_D^w = \rho$. Here, higher ρ can raise trend growth by increasing the growth rate of the data stock, but the magnitude of this effect is controlled by processing: larger π makes changes in retention translate more strongly into idea growth. In particular, along the weak data-processing path, operative data processing and positive retention sustain growth even when $n = 0$, distinguishing the model from standard semi-endogenous specifications.

Data-processing capacity (π, ϕ) plays a *central role* in determining long-run growth. Along any BGP, $g_y = \sigma g_A$ and $g_A = \pi g_D + \phi n$, so (π, ϕ) governs how either source of expanding effective inputs—data accumulation or research-labor growth—translates into idea growth and productivity growth. In particular, if $\pi = \phi = 0$, then $g_A = g_y = 0$ regardless of (ρ, n); but growth can persist with $n = 0$ when $\rho > 0$ (weak regime) or with $\rho = 0$ when $n > 0$ (strong regime).

Figure 1 visualizes these comparative statics by plotting per-capita growth against the processing elasticity π. The weak-processing curve $g_y^w = \sigma(\pi\rho + \phi n)$ rises linearly in π because stronger processing extracts more ideas from a given rate of retention. The strong-processing curve $g_y^s = \sigma\,\frac{n(\pi+\phi)}{1-\sigma\pi}$ steepens with π because stronger processing amplifies the feedback loop. The non-trivial intersection of g_y^w and g_y^s identifies the point at which the economy transitions from the weak data-processing regime to the strong data-processing regime. The upper envelope of g_y^w and g_y^s gives the long-run growth rate at each level of data-processing capacity.

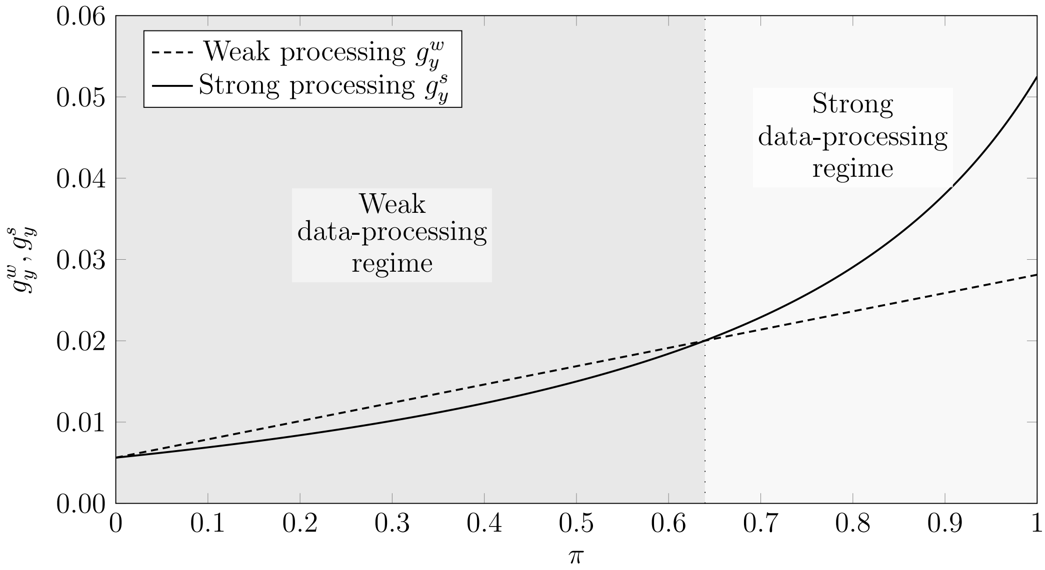

Figure 1: Per-capita output growth as a function of the data-processing elasticity π on the weak and strong data-processing balanced growth paths. Parameterization: $\sigma = 0.75$, $n = 0.01$, $\phi = 0.75$, $\rho = 0.03$.

3 Discussion

This paper proposes a simple way to think about modern growth when ideas are increasingly discovered by working with data. The central distinction is between *accumulating* data and *using* it productively: data becomes economically valuable only when it is processed into implementable ideas. This leads naturally to three technology margins—data processing, data generation, and data retention—with data processing playing the pivotal role because it governs how any expansion in the data stock translates into idea creation and, ultimately, productivity growth.

A key implication is that more data does not automatically mean more growth. Improvements in data generation (more measurement, digitization, and collection) or retention (slower obsolescence, better preservation and usability) can raise the scale of available information, but their macroeconomic payoff depends on whether the economy has enough processing capacity to turn that information into new ideas. When processing capacity is high, the data–ideas–output feedback loop is active and better processing translates into a higher trend growth rate. When processing capacity is low, the economy can accumulate data without proportional productivity gains, with growth being limited by how effectively data is processed and by how persistent usable data is.

The model also suggests a practical interpretation of recent technological change and its policy implications. Many innovations associated with AI and modern data infrastructure can be viewed as improving processing capacity (making it easier to extract ideas from data), retention (keeping data usable over time), or generation (producing more data as a byproduct of activity). The framework clarifies why the same "data revolution" may appear as faster trend growth or simply larger data stocks: the difference is whether processing capacity keeps pace. Accordingly, the highest-leverage policies are those that expand data-processing capacity—skills, research productivity tools, and general-purpose

AI that lowers the cost of turning data into usable ideas. These policies both raise growth directly and increase the growth payoff of improvements in retention and data generation. Without sufficient processing, improvements in retention and generation mainly show up as higher levels and a rising data stock with limited productivity gains.

References

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144.

Cong, L. W., Wei, W., Xie, D., and Zhang, L. (2022). Endogenous growth under multiple uses of data. *Journal of Economic Dynamics and Control*, 141:104395.

Cong, L. W., Xie, D., and Zhang, L. (2021). Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, 67(10):6480–6492.

Farboodi, M. and Veldkamp, L. (2026). A model of the data economy. *Review of Economic Studies*. Forthcoming.

Grossman, G. M. and Helpman, E. (1993). *Innovation and Growth in the Global Economy*. MIT Press.

Jones, C. I. (1995). R&D-based models of economic growth. *Journal of Political Economy*, 103(4):759–784.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858.

Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica*, 65(6):1389–1420.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2):S71–S102.

Segerstrom, P. S. (1998). Endogenous growth without scale effects. *American Economic Review*, 88(5):1290–1310.