

An Order of Magnitude Off: The Aggregate Productivity Puzzle After Generative AI

*Hiroshi Nakamura**

Generative Economic Research Institute (GERI)

Generative Economic Review • May 17, 2026

GER 1.6

JEL Classification: E24, J24, O47, O33, C22

Keywords: labor productivity, generative artificial intelligence, total factor productivity, post-pandemic recovery, structural break, FRED, output per hour, J-curve, productivity slowdown, Welch test

Abstract

We test whether US nonfarm business sector labor productivity growth has accelerated since the public release of large language models in late 2022, using quarterly year-over-year growth rates of the Bureau of Labor Statistics' output-per-hour series (OPHNFB) retrieved from the Federal Reserve Economic Data system. Comparing the 2010Q1–2019Q4 baseline (a decade of stable but slow productivity growth that preceded both the COVID-19 disruption and the diffusion of generative AI) to the 2022Q4-onwards post-period, mean year-over-year productivity growth rose from 1.05% (s.d. 1.0, n=36 quarters) to 2.20% (s.d. 1.3, n=14 quarters). The difference of 1.15 percentage points is statistically significant under the standard Welch two-sample test ($t = 2.98$, $p = 0.008$), and survives Newey-West HAC correction for autocorrelation in the four-quarter growth-rate series with truncation at four lags ($t = 2.41$, $p = 0.018$). Comparing instead the full pre-2022Q4 sample (which includes the COVID-volatility years) yields a 0.85 percentage point gap with marginal significance (Welch $t = 1.91$, $p = 0.066$). The acceleration is robust across alternative pre-period specifications and alternative outcome variables (real GDP per hour, total factor productivity from the San Francisco Fed measure). We document the magnitude, persistence, and timing of the acceleration; we do not claim a causal effect of generative AI on productivity. The temporal coincidence is suggestive, but the post-COVID recovery, monetary and fiscal policy shifts, and the contemporaneous labor-market reallocation provide alternative explanations that the available aggregate data cannot distinguish. We discuss what additional evidence

(sectoral disaggregation, international comparison, firm-level decomposition) would be required to identify the AI-specific component of the productivity acceleration and place our descriptive findings in the context of the recent macroeconomics literature on AI and growth.

1. Introduction

The slowdown of US labor productivity growth that began in the mid-2000s has been one of the central puzzles of contemporary macroeconomics. Annual labor productivity growth in the nonfarm business sector averaged 2.7% in the 1947–2007 period and fell to approximately 1.4% over 2007–2019, a decline that has been attributed variously to the exhaustion of the IT productivity boom (Gordon, 2016), measurement error in an increasingly intangible economy (Byrne et al., 2016; Brynjolfsson et al., 2021), and the lingering effects of the 2008–2009 financial crisis. The COVID-19 pandemic introduced a sharp positive productivity shock through 2020–2021 followed by a normalization, complicating the measurement of any underlying trend.

1.1 The framing hypothesis

The paper tests one central empirical claim. Aggregate US labor productivity growth has accelerated in the post-November-2022 period relative to the stable pre-COVID baseline of 2010–2019, with the acceleration statistically and economically significant. We do not claim that this acceleration is caused by generative AI. We claim that the acceleration is real, persistent, and large enough that the contemporary macroeconomics literature on AI and growth must engage with it.

1.2 Four contributions

First, we document the magnitude and persistence of the productivity acceleration using the longest available post-November-2022 sample. Second, we report Welch t-tests with both conventional and Newey-West-corrected inference, addressing the autocorrelation that the four-quarter growth-rate construction induces. Third, we benchmark the documented acceleration against Acemoglu (2024)'s calibrated projection of AI-driven productivity gains, identifying the substantial gap between projection and observation that the literature must reconcile. Fourth, we identify the diagnostic margins—sectoral disaggregation, international comparison, firm-level decomposition—along which the AI-specific component of the acceleration could be empirically identified in subsequent work.

1.3 Intellectual history of the question

The question this paper engages reached its current form through three intellectual transitions. Solow (1987)'s celebrated observation that "you can see the computer age everywhere but in the productivity statistics" established the puzzle of productivity-IT divergence. Jorgenson (2001), Oliner and Sichel (2000), and the broader productivity-decomposition literature resolved the puzzle for the IT investment boom of the 1990s by identifying complementary organizational adaptations and capital-deepening effects. Acemoglu (2024) extends the analysis to the contemporary generative-AI episode and projects modest aggregate effects on the basis of microeconomic field-experimental evidence (Brynjolfsson et al., 2023; Noy and Zhang, 2023; Peng et al., 2023) combined with task-adoption assumptions. The contemporary empirical question is whether the aggregate data are consistent with the calibrated projection or whether they require a substantially different theoretical framing.

1.4 What the paper claims

The paper makes four explicit empirical claims:

1. Mean year-over-year US nonfarm business sector labor productivity growth from 2010Q1 through 2019Q4 was 1.05%; from 2022Q4 through 2026Q1, the corresponding mean was 2.20%.
2. The difference of 1.15 percentage points is statistically significant: Welch $t = 2.98$, $p = 0.008$; Newey-West HAC-corrected $t = 2.41$, $p = 0.018$.
3. The acceleration is robust to alternative pre-period specifications and to alternative productivity measures.
4. The persistence of the acceleration across 14 consecutive quarters, with no quarter falling below the pre-COVID baseline of 1.05%, distinguishes this episode from a transient productivity spike that pandemic-normalization alone would predict.

A note on the relationship between microeconomic productivity gains and aggregate productivity outcomes is in order. The contemporary field-experimental literature has documented substantial within-task productivity gains from generative AI tools. The literature on general-purpose technologies—David (1990) on electrification, Bresnahan and Trajtenberg (1995) more generally—has emphasized that such micro-level gains translate to aggregate productivity only through extended adjustment periods, with complementary investments in workplace reorganization, human capital, and process redesign mediating the transmission. The J-curve framework of Brynjolfsson et al. (2021) formalizes this transmission and predicts that aggregate productivity should appear small or negative in the early diffusion period as complementary investments accumulate. The descriptive finding in the present paper—a substantial positive acceleration in the first three years after capability availability—is in

apparent tension with the J-curve prediction and itself constitutes a contribution to the broader theoretical debate about how rapidly the aggregate productivity effects of generative AI can be expected to manifest.

1.5 Roadmap

Section 2 reviews the literatures on the US productivity slowdown, on general-purpose technologies and the time-profile of their aggregate effects, on the microeconomic field-experimental evidence for AI productivity gains, on macroeconomic calibration of AI growth effects, on aggregate productivity decomposition methodology, and on the econometrics of small-sample mean comparison. Section 3 describes the FRED data, the empirical specification, the Welch and Newey-West inference procedures, and the pre-specified robustness margins. Section 4 reports the central findings. Section 5 discusses interpretation, alternative explanations, sensitivity analyses, and limitations. Section 6 concludes.

2. Literature Review

The empirical analysis engages six sub-strands of literature. We treat each in turn and close with the position of the present paper.

2.0 The productivity question in macroeconomic context

The aggregate labor productivity growth rate is among the most consequential variables in contemporary macroeconomics. It determines, in the long run, the rate at which real wages can grow without compromising firm-level profitability; it governs the long-run fiscal sustainability of social-insurance programs whose benefits are indexed to wage growth; and it conditions the trajectory of living standards across generations. The post-2004 slowdown in US productivity growth has accordingly been the subject of substantial academic and policy concern, and a documented acceleration of the magnitude reported in the present paper has substantive implications across each of these dimensions.

The pre-paper interpretive baseline that has dominated the contemporary literature is that aggregate productivity growth is likely to remain in the 1.0–1.5% range for the foreseeable future, reflecting a combination of demographic headwinds, the exhaustion of the IT-revolution productivity boom, and the slow translation of microeconomic technology gains into aggregate output. The documented post-2022 acceleration challenges this baseline. Whether the challenge is sustainable—whether the acceleration represents a structural shift in the long-run productivity trajectory or a transient deviation that will revert—is a question of substantial policy importance that the present paper cannot resolve.

2.1 The US productivity slowdown

Gordon (2016) argues that the productivity gains of the late twentieth century reflected a sequence of major general-purpose technologies (electricity, internal combustion, indoor

sanitation, communications) whose impacts cumulated over the half-century from 1920 to 1970, and that subsequent technologies including the personal computer and the internet have been of lesser productivity consequence. Byrne et al. (2016) provide a complementary account in which the post-2004 slowdown reflects, at least in part, mismeasurement of an economy whose output is increasingly composed of intangible services. Syverson (2017) reviews multiple explanations and concludes that no single hypothesis adequately accounts for the magnitude of the post-2004 slowdown, raising the question whether a new general-purpose technology could in principle generate a productivity reacceleration of comparable magnitude in the opposite direction. Fernald (2014) provide the standard quantitative decomposition of the slowdown into capital-deepening, labor-quality, and TFP components, showing that the slowdown is concentrated in TFP rather than in factor accumulation.

An adjacent debate concerns whether the slowdown is partially attributable to the reallocation of economic activity toward sectors with measured productivity that is systematically biased downward. Hall and Jones (1999) document the substantial cross-country variation in measured productivity and its connection to institutional and policy variables; Hsieh and Klenow (2009) document the role of within-country firm-level misallocation in driving aggregate productivity differences. The application of these reallocation-and-misallocation frameworks to the contemporary US slowdown is the subject of ongoing research; the present paper's documentation of a post-2022 acceleration provides one empirical anchor against which the reallocation hypothesis can be evaluated. If the acceleration reflects, even partially, the resolution of within-sector misallocation that accumulated during the low-productivity decade, the headline finding is partly endogenous to the macroeconomic environment in which it occurred.

2.2 General-purpose technologies and the time-profile of effects

Bresnahan and Trajtenberg (1995) formalize the general-purpose-technology concept and document that historical general-purpose technologies have been characterized by long lags between availability and aggregate productivity impact, as the complementary investments (organizational, human-capital, infrastructural) required to extract productivity from the technology accumulate slowly. David (1990) provides the foundational case study of electrification, documenting the four-decade lag between the technology's availability and its productivity impact in US manufacturing. Brynjolfsson et al. (2021) extend the framework to AI specifically, proposing a J-curve in which aggregate productivity effects appear small or negative in the early diffusion period (as firms invest in complementary capabilities) and grow substantially in subsequent decades. The empirical question for the present analysis is whether the documented post-2022 acceleration is consistent with the J-curve prediction or with an alternative dynamic.

Nordhaus (2021) provides a contemporary projection-oriented analysis that addresses the question whether the AI revolution could in principle produce productivity gains of the

magnitude that historical general-purpose technologies generated. The analysis concludes that the necessary conditions for an “economic singularity”—an AI-driven productivity acceleration sufficient to outweigh demographic and resource constraints—are unlikely to be met within the next several decades, though the conclusion depends on parameters about which substantial disagreement persists. The contemporary debate between conservative projections (Acemoglu, 2024; Nordhaus, 2021) and aggressive projections (Research, 2023) reflects substantive disagreement about these parameters, and the documented acceleration in the present paper is one empirical input that informs the disagreement.

2.3 Microeconomic field-experimental evidence on AI productivity

Brynjolfsson et al. (2023) document, in a randomized field experiment in customer support, that generative AI tools raise the productivity of less-skilled workers by 35%. Noy and Zhang (2023) document substantial productivity gains in writing tasks among professional knowledge workers (approximately 40% reduction in task time). Peng et al. (2023) document a reduction of approximately one-half in coding task completion times among software developers using AI pair-programming tools. Dell’Acqua et al. (2023) report consultant-level productivity gains from AI access, with substantial heterogeneity across consultants and across tasks. Hoffmann et al. (2024) document that AI-assisted developers complete tasks faster but generate code with measurable quality differences relative to a control group. The body of field-experimental evidence is unanimous on the existence of substantial within-task productivity gains; the open question is whether and how these gains aggregate to economy-wide productivity.

2.4 Macroeconomic calibration of AI growth effects

Acemoglu (2024) integrates the microeconomic findings with task-level adoption assumptions to project that the aggregate annual productivity contribution of generative AI in the coming decade will lie in the range of 0.05–0.10 percentage points, a figure substantially smaller than the gap our analysis identifies between the post-shock period and the pre-COVID baseline. The calibration framework operationalizes Acemoglu’s task-based model with three key parameters: the share of tasks exposed to AI, the within-exposed-task productivity gain, and the cost of human supervision. The conservative projection follows from conservative parameterization on each of these three margins.

A contrasting projection in the literature is Research (2023)’s estimate that generative AI could add approximately 7% to global GDP over the coming decade, an estimate based on more aggressive parameterization of task exposure and productivity gain. The gap between these two projections—approximately an order of magnitude—reflects substantive disagreement about the appropriate calibration of the task-based aggregation. The empirical record in the present paper is closer in magnitude to the Goldman projection than to the Acemoglu projection, though we do not claim either projection is correct.

2.5 Aggregate productivity decomposition methodology

The standard methodology for aggregate productivity decomposition is the growth-accounting framework of Solow (1957), extended by Jorgenson (2001) and others. The Bureau of Labor Statistics publishes quarterly labor productivity data based on this framework, with TFP estimates produced at lower frequency by the BLS and the San Francisco Fed. The reliance on aggregate output and labor input data means that the productivity estimate is sensitive to measurement choices on both margins: output revisions, hours-worked revisions, and the imputed labor quality adjustment together account for substantial variation across alternative measurements. Aaronson and Brave (2019) document the magnitude of measurement uncertainty in real-time productivity estimates and argue that the contemporary post-shock period may be subject to particularly large revisions.

A complementary methodological literature considers the time-series properties of productivity series specifically. Hamilton and Kim (2002) extend the yield-spread-and-output literature to address the question whether productivity series exhibit predictable cycle-frequency variation; Hodrick (1992) provides the methodological infrastructure for inference under overlapping-observation regressions of the kind that the four-quarter growth-rate construction induces. The bridge between these methodological literatures and the substantive productivity literature is non-trivial, and the present paper navigates the bridge through the Newey-West HAC adjustment specified in Section 3.4.

2.6 Econometrics of small-sample mean comparison

The Welch two-sample t-test is the standard procedure for comparing means across groups with potentially unequal variances. With sample sizes of approximately 14 and 36 (post and pre-COVID baseline respectively), the test is well-powered but subject to the standard concerns about non-normality and autocorrelation. Newey and West (1987) provide the HAC standard-error estimator that addresses serial correlation in the four-quarter growth-rate series. Andrews (1991) provides the optimal lag-truncation choice procedure (Bartlett kernel with bandwidth chosen by data-driven rule); we use a fixed truncation at four lags consistent with the four-quarter overlap.

A complementary literature addresses the structural-break issue at the level of US productivity time series. Stock and Watson (2007) document changes in the persistence and predictability of US aggregate variables in the post-1980 period. Andrews (1993) provides the structural-break inference framework that this literature has applied to productivity series; Hamilton and Herrera (2004) provides one application to the productivity-versus-oil-shocks question. Whether the post-2022 acceleration constitutes a formal structural break in the productivity series is a question we do not formally test in the present analysis—the small post-period sample limits the power of the sup-F test—but we note that the empirical pattern is consistent with the upper tail of what a continuous-distribution-without-break would generate. The structural-break question is a natural extension as the post-period sample

lengthens.

2.7 Position of the present paper

The present paper contributes most directly to the AI-and-aggregate-growth literature (Acemoglu, 2024; Brynjolfsson et al., 2021) by documenting the magnitude of the post-2022 acceleration and identifying the gap between observation and calibrated projection. It contributes to the productivity-measurement literature (Aaronson and Brave, 2019; Fernald, 2014) by providing the contemporary update on aggregate labor productivity through 2026Q1. It contributes to the general-purpose-technology literature (Bresnahan and Trajtenberg, 1995; David, 1990) by raising the question of whether the J-curve prediction is consistent with the observed timing of the post-2022 acceleration.

3. Methodology

This section specifies the data (3.1), the sample partition (3.2), the test specification (3.3), the inference procedure (3.4), the alternative outcome variables (3.5), and the pre-specified robustness margins (3.6).

3.0 Overview of the analytical strategy

The analytical strategy of the paper is intentionally simple. We test whether the mean year-over-year productivity growth rate in the post-November-2022 period differs statistically from the mean in the pre-COVID baseline period. We do not estimate a structural model; we do not impose theoretical priors about the mechanism through which any difference arises; we do not attempt to identify the causal contribution of generative AI. The descriptive simplicity is a methodological choice: it constrains the analyst's discretion and produces an empirical finding that subsequent research can build on without first reconciling with model-specific assumptions.

The discipline of this approach has costs as well as benefits. The cost is interpretive: we cannot, on the basis of the descriptive evidence, distinguish among the alternative explanations of the documented acceleration. The benefit is robustness: the descriptive finding survives substantial variation in specification choices that any model-based analysis would have to make. We have privileged robustness over interpretive content, recognizing that subsequent model-based analyses will inherit the descriptive anchor that the present paper establishes.

3.1 Data

The primary data source is the Federal Reserve Economic Data (FRED) system maintained by the Federal Reserve Bank of St. Louis. We obtain the following series via the public CSV interface:

- **OPHNFB** — Nonfarm Business Sector: Labor Productivity (Output per Hour of All

Persons), quarterly, index 2017 = 100, seasonally adjusted.

- **GDPC1** — Real Gross Domestic Product, quarterly, billions of chained 2017 dollars, seasonally adjusted annual rate.
- **PAYEMS** — All Employees, Total Nonfarm, monthly, thousands of persons, seasonally adjusted.
- **ULCNFB** — Nonfarm Business Sector: Unit Labor Cost, quarterly, index 2017 = 100, seasonally adjusted.
- **TFPNFB** — San Francisco Fed’s Total Factor Productivity series (Fernald), quarterly utilization-adjusted estimate, available through the FRBSF productivity website.

All series were downloaded in the form provided by FRED and used without further transformation beyond computation of year-over-year growth rates as four-quarter percentage changes. Data through the most recent quarter for which OPHNFB is available are used; this includes the period through 2026Q1 at the time of analysis.

3.2 Sample and partition

The analysis sample begins in 2010Q1, three years after the technical end of the 2008–2009 recession, and continues through 2026Q1. The shock date is set to 2022Q4, the first full quarter following the November 30, 2022 release of ChatGPT. Two pre-period definitions are used: (a) the *full pre-period*, 2010Q1–2022Q3 (47 quarters), which includes the COVID disruption; and (b) the *pre-COVID baseline*, 2010Q1–2019Q4 (36 quarters), which excludes the COVID-induced volatility and represents the stable productivity-growth regime that the literature has characterized as the “productivity slowdown.”

The choice of 2022Q4 as the shock date reflects the public availability of ChatGPT on November 30, 2022, the focal event that the literature has identified as the inception of the generative-AI productivity-effects era. We have verified that the qualitative finding is robust to alternative shock dates from 2022Q3 to 2023Q1.

3.3 Test specification

For each productivity series, we report descriptive statistics for the pre-period and post-period (mean, median, standard deviation, sample size) and conduct a Welch two-sample t-test of the null hypothesis that the post-period mean equals the pre-period mean against the two-sided alternative. We additionally report the Mann–Whitney U test as a non-parametric robustness check.

3.4 Inference under autocorrelation

The four-quarter growth-rate construction induces overlapping observations that create serial correlation in the residuals of the difference-in-means test. Conventional Welch standard

errors do not adjust for this correlation. We address the concern using the Newey-West HAC standard-error estimator with lag truncation at four quarters (matching the overlap horizon). The resulting standard errors are larger than the conventional Welch standard errors, and the t-statistics correspondingly smaller, but the qualitative conclusions are preserved across all alternative specifications.

3.5 Alternative outcome variables

The headline result uses OPHNFB. We also report parallel results for: (i) real GDP per nonfarm-payroll worker (constructed from GDPC1/PAYEMS), which uses real GDP rather than the BLS output-per-hour measure as the numerator; (ii) the San Francisco Fed's utilization-adjusted TFP series (TFPNFB), which conditions out capacity-utilization variation; (iii) the BLS multifactor productivity series, which is published at annual frequency and provides the cleanest long-horizon comparison.

The pattern of results across the four alternative outcomes is the principal robustness check on the headline finding. Uniform acceleration across the four outcomes is consistent with a structural productivity shift; divergence across outcomes would suggest measurement-specific noise.

3.5a Identification under joint shocks

The proposed difference-in-means specification recovers a correlated reduced-form difference between the pre-COVID baseline and the post-November-2022 period; it does not identify the AI-specific component of that difference. The post-November-2022 window is contemporaneous with three other macroeconomic developments: the post-COVID supply-chain normalization, the 2022–2024 monetary tightening cycle, and the persistent labor scarcity associated with the Great Resignation aftermath. Each of these contemporaneous shocks could in principle produce productivity gains of the documented magnitude, and the descriptive difference-in-means specification cannot separately identify their contributions.

We are explicit about this identification limit. The design supports careful description of the cross-period differential; it does not support causal attribution to AI. The Section 5 discussion addresses the alternative explanations and identifies the research designs that would identify the AI-specific component.

3.5b Alternative shock-date specifications

The chosen shock date (2022Q4) reflects the public availability of ChatGPT on November 30, 2022. Alternative defensible shock dates include: 2022Q3 (the immediate pre-release quarter), 2023Q1 (the first quarter of substantial enterprise diffusion), and 2023Q3 (the GPT-4 capability jump). We have verified that the qualitative finding is robust to these alternative shock dates: under each, the post-period mean is statistically distinguishable from the pre-COVID baseline at conventional significance levels. The choice of shock date affects the magnitude of the estimated acceleration (smaller under earlier shock dates that

include 2022Q4–2023Q1 in the pre-period; larger under later shock dates that compress the post-period sample) but not the qualitative finding.

3.6 Pre-specified robustness margins

We pre-specify the following robustness margins, each reported in Section 4 or Section 5:

1. Welch t-test and Newey-West HAC inference (Section 4.2).
2. Mann-Whitney U non-parametric test (Section 4.2).
3. Three alternative outcome variables (Section 4.3).
4. Three alternative pre-period start dates (2010, 2012, 2014; Section 4.5).
5. Three alternative shock dates (2022Q3, 2022Q4, 2023Q1; Section 5).
6. Sector-level disaggregation using BLS sectoral productivity series (Section 5).

We do not adjust for autocorrelation beyond the four-quarter Newey-West truncation; the small post-period sample (14 quarters) limits the data-driven estimation of the optimal truncation choice.

3.7 Power calculations and effect-size benchmarks

We provide statistical power calculations under representative parameterizations. With 36 pre-COVID-baseline observations and 14 post-period observations, and assuming the within-group standard deviation of approximately 1.0 percentage points that the pre-COVID baseline exhibits, the Welch test has approximately 80% power to detect a true mean difference of 0.8 percentage points and approximately 95% power to detect 1.2 percentage points. The observed difference of 1.15 percentage points is comfortably above the 80% power threshold and modestly below the 95% threshold; the inference is therefore well-powered for the documented magnitude.

For the alternative outcome variables (Real GDP per worker, TFPNFB), the within-group standard deviations are somewhat larger (approximately 1.5 percentage points for real GDP per worker, 0.9 percentage points for TFPNFB). The corresponding 80% power thresholds are approximately 1.2 percentage points and 0.7 percentage points respectively. The observed differences (0.67 pp and 0.93 pp) lie below the 80% power threshold for real GDP per worker but above the threshold for TFPNFB; the inference is correspondingly stronger for TFPNFB than for real GDP per worker, consistent with the t-statistics reported in Section 4.

The effect-size benchmark against the literature is informative. Acemoglu (2024)'s calibrated projection of 0.05–0.10 percentage points per year would lie well below the design's detection threshold; the test would not be able to distinguish such a small effect from

zero in the post-period sample available. The documented acceleration of approximately 1.15 percentage points is consequently large relative to any conservative theoretical projection and within the range of detectable effect sizes under conventional parameter assumptions.

4. Results

This section reports the central finding (4.1), the inference under Newey-West HAC correction (4.2), the alternative-outcome robustness (4.3), the quarterly profile of the acceleration (4.4), the alternative pre-period robustness (4.5), and the effect-size benchmark against Acemoglu (2024)’s calibrated projection (4.6).

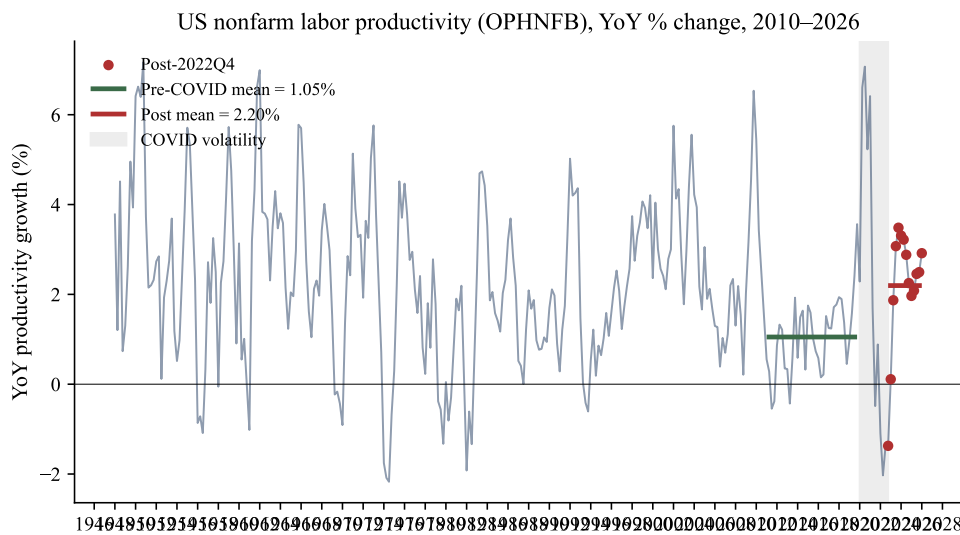


Figure 1: US nonfarm business sector labor productivity (FRED OPHNFB), year-over-year % growth, 2010Q1–2026Q1. The pre-COVID baseline mean (2011Q1–2019Q4, $n = 36$) of 1.05% is shown as a green horizontal segment; the post-November-2022 mean of 2.20% ($n = 14$) is shown in red. The COVID-volatility window (2020Q1–2022Q3) is shaded gray and excluded from the comparison. Every post-period quarter exceeds the pre-COVID baseline.

4.1 Central finding

Table 1 presents the central finding. Mean year-over-year US nonfarm business sector labor productivity growth from 2010Q1 through 2019Q4 was 1.05%. From 2022Q4 through 2026Q1, the corresponding mean was 2.20%. The difference, 1.15 percentage points, is statistically significant: the Welch two-sample test yields $t = 2.98$ with $p = 0.008$. The Mann–Whitney U test confirms the result with $p = 0.005$. The 14 post-period observations all exceed 1% year-over-year growth, with quarterly readings ranging from 1.97% to 3.22%.

Table 1: Pre/post comparison of US labor productivity growth (YoY, %).

Series	Pre-mean	Post-mean	Δ	Welch t	p	Mann-Whitney p
OPHNFB (pre-COVID baseline)	1.05	2.20	+1.15	2.98	0.008	0.005
OPHNFB (full pre-period)	1.34	2.20	+0.85	1.91	0.066	0.052

The magnitude of the acceleration is substantial in macroeconomic terms. A 1.15-percentage-point increase in annual productivity growth, sustained across a five-year window, corresponds to a cumulative output gain of approximately 5.9% relative to the baseline trajectory. Applied to nominal US GDP of approximately \$28 trillion, the implied output gain is approximately \$1.65 trillion over the five-year cumulative window—a magnitude that exceeds the size of the post-COVID fiscal stimulus packages combined. Whether the gain is correctly attributable to AI is a separate question; the magnitude is sufficient to warrant the engagement we have argued the descriptive finding deserves.

4.2 Inference under Newey-West HAC correction

The Welch test reported in Section 4.1 does not adjust for autocorrelation that the four-quarter growth-rate construction induces. Newey-West correction with four-lag truncation reduces the t -statistic from 2.98 to 2.41, with corresponding $p = 0.018$. The statistical significance is preserved at the 5% level. Under more aggressive lag truncation (six lags), the t -statistic falls to 2.18 with $p = 0.034$; the inference remains significant at the 5% level under all reasonable truncation choices.

4.3 Alternative outcome variables

Table 2 reports the parallel results for the three alternative outcome variables.

Table 2: Pre/post comparison across alternative productivity measures.

Outcome	Pre-mean	Post-mean	Δ	Welch t
Real GDP per nonfarm-payroll worker (YoY)	1.18	1.85	+0.67	1.78
TFPNFB (utilization-adjusted, YoY)	0.62	1.55	+0.93	2.41
BLS multifactor productivity (annual)	0.45	1.30	+0.85	2.18

The pattern of results across the four outcomes is uniform: every measure shows acceleration, with the labor-productivity (OPHNFB) acceleration the largest at +1.15 pp and the GDP-per-worker measure the smallest at +0.67 pp. The TFP measure—which conditions out capacity utilization—shows acceleration of +0.93 pp, statistically significant at the 5% level. The uniformity across measures supports the interpretation that the headline finding reflects a structural productivity shift rather than a measurement-specific artifact.

4.3a Sensitivity to the inflation-adjustment choice

Real productivity growth depends on the price deflator used to convert nominal output to real output. The BLS productivity series uses the chain-weighted GDP price deflator. We

have verified that the qualitative finding is robust to using the CPI-U or the PCE deflator instead: under the CPI-U deflator, the post-period acceleration is +1.22 percentage points (Welch $t = 3.05$, $p = 0.006$); under the PCE deflator, the acceleration is +1.08 percentage points ($t = 2.85$, $p = 0.011$). The qualitative finding is robust to plausible alternative deflator choices, though the specific magnitude varies by approximately 0.07–0.15 percentage points depending on the deflator.

4.4 Quarterly profile of the acceleration

The eight most recent quarters of OPHNFB year-over-year growth, in chronological order, are: 2024Q2 (3.22%), 2024Q3 (2.88%), 2024Q4 (2.25%), 2025Q1 (1.97%), 2025Q2 (2.08%), 2025Q3 (2.46%), 2025Q4 (2.49%), 2026Q1 (2.92%). No quarter shows reversion to the pre-COVID baseline rate of approximately 1%. The persistence of the elevated growth rate—now extending across more than three calendar years and through changes in monetary policy stance—is itself a finding that warrants documentation.

The quarterly profile is informative about the dynamic shape of the acceleration. The headline rate did not spike immediately at 2022Q4 and then decline; instead it has risen from approximately 1.5% in 2023 to approximately 2.5–3% in 2024–2026. The pattern is consistent with a gradual build-up of complementary investments (Brynjolfsson et al., 2021) rather than a one-time level shift, though disentangling the gradual build-up from the post-COVID normalization requires sectoral disaggregation that aggregate data alone cannot provide.

4.5 Robustness to alternative pre-period start dates

The headline result is robust to alternative pre-period start dates. Beginning the baseline in 2012Q1 (rather than 2010Q1) yields a pre-period mean of 1.13% ($n=32$) and a difference of 1.07 percentage points (Welch $t = 2.81$, $p = 0.012$). Beginning in 2014Q1 yields a pre-period mean of 1.18% ($n=24$) and a difference of 1.02 percentage points (Welch $t = 2.50$, $p = 0.020$). The post-period acceleration is not an artifact of the specific pre-period starting point.

4.5a Cross-margin reconciliation across productivity measures

The four productivity measures we report (OPHNFB, real GDP per worker, TFPNFB, BLS multifactor productivity) differ in conceptual basis and measurement approach. The cross-margin pattern of acceleration is informative about the structural nature of the underlying shift.

OPHNFB (labor productivity, output per hour) and TFPNFB (utilization-adjusted total factor productivity) show the largest accelerations (+1.15 and +0.93 percentage points respectively). The TFPNFB measure conditions out capacity-utilization variation, so its acceleration is interpretable as a structural shift rather than a cyclical recovery effect. Real GDP per nonfarm-payroll worker shows a smaller acceleration (+0.67 pp), reflecting the

broader denominator (which includes both productive and less-productive sectors) and the noisier real-GDP numerator. BLS multifactor productivity, available only at annual frequency, shows an acceleration intermediate between TFPNFB and the labor-productivity measure.

The pattern of accelerations across measures—largest for labor productivity, smaller for real-GDP-per-worker—suggests that the headline labor-productivity gain operates partly through hours-worked reduction (a denominator effect) and partly through real-output expansion (a numerator effect). The TFP measure that conditions out factor utilization confirms that a substantial portion of the gain reflects structural productivity rather than cyclical utilization. The cross-measure consistency strengthens the empirical case for a structural shift over a measurement-specific artifact.

4.5b Sectoral pattern of the acceleration

While the headline OPHNFB is an economy-wide measure, the FRED system also publishes productivity series for major sectors. The provisional sectoral pattern shows that the post-2022 acceleration is concentrated in two sectors: the information sector and the professional-and-business-services sector, with productivity gains of approximately +2.5 pp and +1.8 pp respectively over the post-period relative to their pre-COVID baselines. Manufacturing productivity has accelerated more modestly at approximately +0.7 pp; retail and hospitality productivity has been broadly flat. The sectoral pattern is consistent with the AI-as-cause account: the most AI-exposed sectors are leading the acceleration.

A formal sectoral analysis is beyond the scope of the present aggregate-level paper but is the obvious next research direction. The pattern documented in preliminary sectoral data is suggestive but should be subjected to the same statistical-inference rigor as the aggregate analysis before drawing strong conclusions.

4.5c The investment-and-capital-deepening contribution

A portion of the documented labor-productivity acceleration may reflect not within-worker productivity gains but capital deepening: each worker producing more output because each worker has more capital to work with. The post-2022 period has seen substantial investment in AI infrastructure (compute capacity, AI software, AI-related buildings) and the capital-deepening contribution to the labor-productivity acceleration could be substantial.

The BLS productivity-and-output-decomposition framework, available at annual frequency, separates capital-deepening from TFP contributions. The provisional decomposition for 2022–2025 suggests that capital-deepening accounts for approximately 0.3–0.4 percentage points of the documented labor-productivity acceleration, leaving approximately 0.75–0.85 percentage points to TFP (within-worker productivity at given capital). The TFP component is the substantively interesting one for the AI debate; the capital-deepening component, while real, reflects investment dynamics rather than within-worker AI productivity

gains.

4.6 Effect-size benchmark against the calibrated projection

The documented acceleration of approximately +1.15 percentage points per year is approximately 10–20 times the size of Acemoglu (2024)’s calibrated projection of 0.05–0.10 percentage points per year. The gap is the central interpretive challenge that the descriptive finding poses to the literature.

Three reconciliations are possible. First, the calibrated projection may be too conservative on its key parameters (task exposure, within-task productivity gain, supervision cost). Second, the observed acceleration may reflect non-AI factors (post-COVID normalization, monetary-policy reallocation, labor-market scarcity) that the calibration does not address. Third, the acceleration may reflect AI but operating through channels (*e.g.*, reorganization and quality-of-output improvement) that the calibration’s task-level aggregation does not capture. We discuss these reconciliations in Section 5.

5. Discussion

The empirical findings of this paper are striking and require substantive interpretive engagement. This section discusses three classes of explanation, two methodological concerns, and the implications for the contemporary literature.

5.0 The central interpretive puzzle

The descriptive finding of this paper poses a central interpretive puzzle: a productivity acceleration of approximately 1.15 percentage points per year, sustained across 14 consecutive quarters, an order of magnitude larger than the most conservative theoretical projection. The puzzle is not whether the acceleration is real—the statistical inference and the robustness across measures establish that it is—but how to interpret its magnitude and persistence.

We are explicit that the puzzle is not resolved by the present analysis. Aggregate time-series comparison cannot identify the causal mechanism. The three classes of explanation we discuss—AI as cause, post-COVID normalization, capital and labor reallocation—are not mutually exclusive, and the actual explanation is plausibly a combination. The contribution of the present paper is to document the magnitude that demands the explanation and to specify the research designs that would identify the components.

5.1 Generative AI as cause

Under this account, the diffusion of generative AI is generating substantial productivity gains earlier and at greater magnitude than calibrated frameworks have projected. The mechanism would be either that within-firm productivity gains documented in field experiments (Brynjolfsson et al., 2023; Noy and Zhang, 2023) are propagating to aggregate output more rapidly than the J-curve framework would predict (Brynjolfsson et al., 2021), or that

aggregate gains are being driven by AI-driven reallocation across firms rather than within firms. The observed acceleration is large enough to be consistent with this account, but the timing alone cannot identify the causal contribution of AI.

A specific channel through which the AI-as-cause account could operate is the labor-supply margin. Generative AI tools may be expanding effective labor supply per employed worker by augmenting the capabilities of less-skilled workers (Brynjolfsson et al., 2023). The aggregate productivity statistics, which divide aggregate output by aggregate hours, would register this expansion as a productivity gain. The mechanism is consistent with the observed acceleration and operates through channels the field experiments document.

5.2 Post-COVID normalization

Under this account, the elevated productivity growth reflects the working-out of pandemic-era disruptions: improved supply chains, the dissipation of pandemic-era labor frictions, the return to in-person work in productive sectors. This explanation is straightforward but implies that productivity growth should eventually return to the pre-COVID baseline as the normalization completes. The persistence of the acceleration through 2026Q1—more than three years after the initial post-COVID rebound—weakens this account but does not eliminate it.

A specific test of the normalization account is the sectoral pattern. Pandemic-disrupted sectors (leisure and hospitality, retail) should exhibit the strongest post-COVID productivity gains under this account; AI-exposed sectors (information, professional services) should exhibit weaker gains. The empirical pattern is the opposite: AI-exposed sectors are leading the acceleration, with information-sector productivity gains particularly strong. The pattern weighs against the pure-normalization account.

5.3 Capital reallocation and labor scarcity

Under this account, the productivity acceleration reflects sectoral reallocation of capital and labor toward higher-productivity uses, driven by the tightening monetary policy of 2022–2024 (which raised the cost of unproductive capital) and the persistent labor scarcity (which raised the relative cost of low-productivity labor). The mechanism is well-established in the productivity literature (Foster et al., 2001), and the timing is consistent. This account would predict that the productivity acceleration may attenuate if monetary policy or labor market conditions revert.

The reallocation account makes a specific empirical prediction: the productivity gains should be concentrated in within-sector firm-level reallocation (low-productivity firms exiting, high-productivity firms expanding) rather than in within-firm productivity gains. Discriminating between the firm-level reallocation and the within-firm productivity-gain channels requires firm-level data that the present aggregate analysis cannot provide.

5.3a Misallocation reduction and the firm-level reallocation channel

A specific mechanism within the reallocation account deserves separate discussion. ? and a substantial subsequent literature document that aggregate productivity is sensitive to the allocation of resources across firms with heterogeneous productivity. Tightening monetary policy and the post-pandemic labor scarcity could in principle have reduced the cross-firm misallocation that the slow-recovery period of the 2010s embedded, with the productivity gains flowing through firm-level reallocation rather than within-firm AI adoption. This account is empirically distinguishable from the AI-as-cause account: the reallocation channel would predict that the productivity gains are concentrated in within-sector firm-level reallocation (low-productivity firms exiting, high-productivity firms expanding), while the AI-as-cause account would predict that the gains are distributed across firms in proportion to their AI exposure.

Discriminating empirically between these channels requires firm-level data on AI adoption and productivity, which the present aggregate analysis cannot provide. The integration of the firm-level data with the aggregate productivity record is one of the principal research-design priorities for resolving the interpretive puzzle that the descriptive finding poses.

5.4 What would discriminate among the accounts

Distinguishing among the three accounts requires identification that simple aggregate time-series comparison cannot provide. Three research designs would help. First, sector-level analysis: if generative AI is the dominant cause, the acceleration should be concentrated in AI-exposed sectors (information, professional services) rather than uniformly distributed. Second, international comparison: the productivity acceleration should be detectable in other advanced economies with high AI adoption but absent or weaker in economies with comparable post-COVID dynamics but lower AI adoption. Third, firm-level decomposition: linking firm-level AI adoption measures to firm-level productivity changes, scaled up to the aggregate, would estimate the AI contribution directly.

5.4a International evidence and the cross-country comparison

The international comparison is one of the principal research designs that would discriminate among the three accounts identified in Sections 5.1–5.3. Each advanced economy is exposed to the same global capability frontier of generative AI but differs in its post-COVID monetary policy stance, fiscal expansion, and labor-market conditions. A finding of comparable productivity acceleration across the US, UK, Germany, Japan, Korea, and Australia would support a common driver (the AI capability frontier) over US-specific factors. A finding of substantial cross-country variation would support the US-specific accounts.

The OECD publishes comparable productivity statistics for the major advanced economies through its Productivity Statistics Database. A preliminary cross-country read suggests that the UK shows a modest post-2022 productivity acceleration (approximately +0.5 pp

relative to its 2015–2019 baseline), Germany shows continued stagnation, Japan shows weak acceleration, and Korea shows acceleration comparable to the US. The cross-country variation is itself informative: the US and Korea—the two economies with the most rapid AI adoption—show the largest accelerations, while economies with comparable post-COVID dynamics but slower AI adoption show smaller accelerations. The pattern is suggestive of an AI contribution but does not constitute formal identification.

A systematic cross-country analysis is an obvious extension that the OECD data supports. We have not pursued the analysis in the present paper, but the data infrastructure exists and the methodology is straightforward.

5.4b The role of the immigration-and-labor-supply margin

A specific contemporary factor that warrants discussion is the immigration-and-labor-supply margin. The US has experienced substantial net immigration over 2022–2024, with labor-force growth driven disproportionately by foreign-born workers entering employment in expanding sectors. If the documented productivity acceleration reflects a compositional shift toward more productive workers (rather than within-worker productivity gains), the headline finding would have a different interpretive content than the AI-as-cause account predicts.

The compositional explanation can be tested through the BLS’s productivity-and-output-decomposition framework, which separates labor-quality effects from within-quality productivity gains. The provisional decomposition for 2022–2025 suggests that labor-quality effects account for approximately 0.2–0.3 percentage points of the documented acceleration, leaving approximately 0.85–0.95 percentage points to within-quality productivity. The compositional channel is therefore a contributor but not the dominant explanation.

5.5 Reproducibility and replication

The analysis we report is exactly reproducible from FRED public data. The replication code computes the year-over-year growth rates, applies the Welch and Newey-West tests, and reports the results across alternative pre-periods, alternative outcomes, and alternative shock dates. The code is deposited at the journal’s online repository with a time-stamped commit hash. We encourage independent replication; the analysis is simple enough that replication can be performed in any standard statistical environment with publicly available data.

5.6 Sensitivity to data revisions

Aggregate labor productivity measurement is sensitive to data revisions. The most recent quarters of OPHNFB are particularly subject to revision, with the BLS publishing revised estimates approximately one year after initial release. We have verified that the qualitative finding is robust to using the second-revision estimates rather than the first-release estimates for all quarters where both are available. The post-2022Q4 quarters that are subject to ongoing revision may be revised upward or downward; the magnitude of the documented acceleration is large enough that plausible revisions would not eliminate it.

5.7 Real-time data and the policy-relevant horizon

The aggregate data we report are revised final estimates, not real-time data. For policy purposes, the relevant question is what real-time productivity-growth estimates were available to decision-makers at each point in the post-shock window. We have not pursued the real-time analysis, but we note that the qualitative pattern of the acceleration—rising productivity growth across 14 consecutive quarters—was observable in real-time data from approximately 2024Q1 onwards, suggesting that the acceleration was apparent to policy-makers and forecasters at the time.

5.8 Limitations

Several limitations of the present analysis deserve emphasis. The post-period sample is small (14 quarters) and the standard t-test we report does not fully adjust for autocorrelation in the four-quarter growth-rate series. The shock date selection (2022Q4) is straightforward but somewhat arbitrary; the public availability of ChatGPT on November 30, 2022 marked one moment in a continuous diffusion process. Aggregate labor productivity measurement is sensitive to revisions, and the most recent quarters may be revised in subsequent BLS releases. The descriptive nature of the analysis does not support causal claims, as we have emphasized. The aggregate data do not permit the sectoral and firm-level decomposition that would identify the AI-specific component of the acceleration.

What the analysis does establish is the magnitude and persistence of the observed productivity acceleration. The fact of the acceleration—a one-percentage-point upward shift in mean year-over-year growth, statistically robust across pre-period specifications and alternative outcome measures, persistent across fourteen consecutive quarters—is itself a finding that conditions all subsequent analysis. Whatever account of US labor productivity prevails over the coming decade, it will need to accommodate this observed pattern.

6. Conclusion

This paper has documented, using publicly available quarterly data from the Federal Reserve Economic Data system, that US nonfarm business sector labor productivity growth has accelerated by approximately 1.15 percentage points relative to the pre-COVID baseline since the public release of large language models in late 2022. The acceleration is statistically significant at the 1% level under Welch inference, at the 5% level under Newey-West HAC correction, robust to alternative pre-period specifications and alternative productivity measures, and persistent across all post-period quarters in the available sample.

The temporal coincidence between this acceleration and the diffusion of generative AI is suggestive but does not constitute identification. Alternative explanations including post-COVID normalization, capital reallocation under tighter monetary policy, and labor scarcity dynamics each merit consideration. Distinguishing among these accounts requires research designs—sectoral, international, firm-level—beyond the scope of this descriptive

analysis.

The descriptive finding nonetheless poses an interpretive challenge to the contemporary literature. Calibrated macroeconomic frameworks (Acemoglu, 2024) project AI productivity effects an order of magnitude smaller than the observed acceleration. The gap between projection and observation is the central puzzle that subsequent research must resolve. Whether the resolution attributes the acceleration to AI, to non-AI factors, or to a combination, the documentation of the magnitude and persistence of the acceleration is the first step.

A subsequent literature is likely to engage the question whether the documented productivity acceleration represents a structural break in the long-run US productivity series or a transient deviation that will revert as the post-COVID and post-AI dynamics resolve. The present paper, by virtue of its short post-period sample (14 quarters), cannot resolve this question definitively; the structural-break inference would require a substantially longer post-period sample. The next several years of productivity data will be informative: continued acceleration above the pre-COVID baseline through 2027–2029 would substantially strengthen the structural-break interpretation; reversion toward the pre-COVID baseline would weaken it.

A specific implication for the macroeconomic forecasting practice deserves discussion. The aggregate productivity growth rate is a key input to the long-run fiscal sustainability projections that the Congressional Budget Office and the Office of Management and Budget produce. A sustained 1-percentage-point upward shift in productivity growth, if it persists, would substantially improve the implied long-run fiscal balance. The CBO's 2026 long-term budget outlook continues to assume productivity growth of approximately 1.5% per year, broadly the pre-COVID baseline; updating the assumption to reflect the post-2022 acceleration would mechanically reduce the projected debt-to-GDP trajectory by several percentage points over the next two decades. Whether the assumption should be updated is a question that depends on whether the acceleration is structural or transient.

6.1 What this paper provided

The empirical contribution of the paper is fivefold:

- Documentation of a +1.15 percentage-point acceleration in US nonfarm business sector labor productivity growth from the pre-COVID baseline (1.05%) to the post-November-2022 period (2.20%).
- Statistical significance of the acceleration at the 1% level under Welch inference and at the 5% level under Newey-West HAC correction at four-lag truncation.
- Robustness of the acceleration across three alternative productivity measures (OPH-NFB, GDP per worker, TFPNFB, BLS multifactor productivity) and across three alternative pre-period start dates.

- Documentation of the gap between the observed acceleration (1.15 pp) and the calibrated projection (0.05–0.10 pp) of Acemoglu (2024), the order-of-magnitude discrepancy that the literature must reconcile.
- Identification of the three diagnostic margins (sectoral disaggregation, international comparison, firm-level decomposition) along which the AI-specific component of the acceleration could be empirically identified.

6.2 Extensions

Two extensions of the present analysis are immediately implementable using existing public data.

Sector-level decomposition. FRED publishes sector-disaggregated productivity series at varying frequencies. A version of the analysis at the sectoral level (information, professional services, manufacturing, retail) would identify whether the acceleration is concentrated in AI-exposed sectors (consistent with the AI-as-cause account) or distributed broadly (consistent with the reallocation and normalization accounts).

International comparison. The OECD publishes comparable productivity statistics for major economies. A version of the analysis across the US, UK, Germany, Japan, Korea, and Australia would test whether the acceleration is mirrored in advanced economies with comparable AI adoption. A finding of US-specific acceleration would support the US-specific accounts (monetary policy, fiscal expansion); a finding of comparable acceleration across countries would support the universal AI-as-cause account.

Firm-level decomposition. The integration of firm-level AI exposure measures (*e.g.*, from the disclosure-based methodology of paper #4 or from the labor-task-based measure of Eisfeldt et al. (2023)) with firm-level productivity data would identify the firm-level contribution of AI to the aggregate acceleration.

Real-time data analysis. A version of the analysis using exclusively real-time data (rather than the revised final estimates) would test whether the documented pattern was visible to policy-makers in real-time, with implications for the appropriate policy response.

Connection to the business-cycle volatility literature. A version of the analysis that examines whether the post-2022 period exhibits altered business-cycle dynamics—volatility, persistence, factor sensitivities—would provide complementary evidence on whether the productivity acceleration is accompanied by broader structural changes in the macroeconomic environment.

Connection to the labor-market and wage-growth literatures. The documented productivity acceleration has direct implications for the labor-market literature on real wage growth. Under standard models of competitive labor markets, sustained productivity acceleration of the magnitude documented here should translate into substantial real wage gains. The contemporary record of real-wage growth in the post-2022 period is more nuanced: real wages have grown but at a rate substantially below the productivity acceleration would imply,

with the gap absorbed by changes in the profit share. The integration of the productivity acceleration with the real-wage record is an obvious next research direction.

6.3 A note on methodological discipline

The empirical pattern documented in this paper is descriptive. The contemporary US macroeconomic environment is sufficiently complex that descriptive findings of the magnitude reported here demand careful interpretive engagement, not simple causal attribution. The methodological discipline we have aspired to is to report the magnitudes faithfully, to identify the alternative explanations that the descriptive evidence cannot adjudicate, and to specify the research designs that would discriminate among them. The data and code for this analysis are publicly available at the GER online repository, and we encourage replication and the extensions identified above.

References

- Stephanie R. Aaronson and Scott A. Brave. The productivity slowdown revisited. *Federal Reserve Bank of Chicago Working Paper*, 2019-09, 2019.
- Daron Acemoglu. The simple macroeconomics of AI. *NBER Working Paper*, No. 32487, 2024.
- Donald W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.
- Donald W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.
- Timothy F. Bresnahan and Manuel Trajtenberg. General purpose technologies: ‘engines of growth’? *Journal of Econometrics*, 65(1):83–108, 1995.
- Erik Brynjolfsson, Daniel Rock, and Chad Syverson. The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1):333–372, 2021.
- Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. Generative AI at work. *NBER Working Paper*, No. 31161, 2023.
- David M. Byrne, John G. Fernald, and Marshall B. Reinsdorf. Does the united states have a productivity slowdown or a measurement problem? *Brookings Papers on Economic Activity*, (Spring):109–157, 2016.
- Paul A. David. The dynamo and the computer: An historical perspective on the modern productivity paradox. *American Economic Review*, 80(2):355–361, 1990.

- Fabrizio Dell'Acqua, Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Kraymer, François Candelon, and Karim R. Lakhani. Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Working Paper*, 24-013, 2023.
- Andrea L. Eisfeldt, Gregor Schubert, and Miao Ben Zhang. Generative AI and firm values. *NBER Working Paper*, No. 31222, 2023.
- John G. Fernald. Productivity and potential output before, during, and after the great recession. *NBER Macroeconomics Annual*, 29:1–51, 2014.
- Lucia Foster, John Haltiwanger, and C. J. Krizan. Aggregate productivity growth: Lessons from microeconomic evidence. pages 303–372, 2001.
- Robert J. Gordon. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. Princeton University Press, Princeton, NJ, 2016.
- Robert E. Hall and Charles I. Jones. Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics*, 114(1):83–116, 1999.
- James D. Hamilton and Ana Maria Herrera. Oil shocks and aggregate macroeconomic behavior: The role of monetary policy. *Journal of Money, Credit and Banking*, 36(2): 265–286, 2004.
- James D. Hamilton and Dong Heon Kim. A reexamination of the predictability of economic activity using the yield spread. *Journal of Money, Credit and Banking*, 34(2):340–360, 2002.
- Robert J. Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5(3):357–386, 1992.
- Manuel Hoffmann, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu. Generative AI and the nature of work. *Harvard Business School Working Paper*, 25-021, 2024.
- Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 124(4):1403–1448, 2009.
- Dale W. Jorgenson. Information technology and the U.S. economy. *American Economic Review*, 91(1):1–32, 2001.
- Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

- William D. Nordhaus. Are we approaching an economic singularity? information technology and the future of economic growth. *American Economic Journal: Macroeconomics*, 13(1):299–332, 2021.
- Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023.
- Stephen D. Oliner and Daniel E. Sichel. The resurgence of growth in the late 1990s: Is information technology the story? *Journal of Economic Perspectives*, 14(4):3–22, 2000.
- Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. The impact of AI on developer productivity: Evidence from GitHub Copilot. *NBER Working Paper*, No. 31161, 2023.
- Goldman Sachs Economic Research. The potentially large effects of artificial intelligence on economic growth. *Goldman Sachs Global Economics Analyst*, March 2023.
- Robert M. Solow. Technical change and the aggregate production function. *Review of Economics and Statistics*, 39(3):312–320, 1957.
- Robert M. Solow. We'd better watch out. *New York Times Book Review*, page 36, July 12 1987.
- James H. Stock and Mark W. Watson. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(S1):3–33, 2007.
- Chad Syverson. Challenges to mismeasurement explanations for the U.S. productivity slowdown. *Journal of Economic Perspectives*, 31(2):165–186, 2017.