Measuring the Impact of AI on Productivity: Cross-Sectoral Evidence from the U.S. and Eighteen Other Economies Post-ChatGPT

Benjamin Verschuere[1] Angus Cameron[2]

April 27, 2026. Current version: June 4, 2026.[3]

Abstract

We provide the first cross-country sectoral measurement of artificial intelligence's contribution to trend productivity growth globally, identified from observed sectoral outcomes through 2025Q4 rather than from forecasts of diffusion. Using a panel difference-in-differences specification with AI exposure measured by the Workflow Completeness construct (C^*) of Verschuere and Cameron (2026a), we estimate that AI is contributing approximately $1.26 trillion per year across the United States, the EU 12, and China combined at the end-of-2025 run-rate. The U.S. accounts for $878 billion (95 percent CI [$602B, $1,155B]) across five primary sectors covering roughly 54 percent of GDP, raising aggregate growth on this base to +5.0 percent per year against a +2.2 percent counterfactual. The EU 12 pooled panel contributes $138 billion (95 percent CI [$23B, $254B]) and China $240 billion (95 percent CI [$116B, $365B]). The result holds under the primary inference cascade (wild cluster bootstrap and permutation), falls short of significance only under its single most conservative bootstrap variant, and survives ten alternative-mechanism tests (eight conducted in this paper, two adopted from the companion papers).

Keywords: AI productivity, ChatGPT, sectoral productivity, cross-country panel, monetary policy.
JEL codes: O33, E22, E32, O47.

Full disclaimer and conflict of interest disclosure appear on page 2.

[1]Liminal Capital LLC. ben.verschuere@liminal-capital.com
[2]Liminal Capital LLC. angus.cameron@liminal-capital.com

[3]Changes since the original draft: deepened the C^* construction to the full O*NET task corpus (from a previously narrower one); re-pinned all reported estimates to a single canonical specification for consistency; and corrected the China attribution to the end-of-2025 run-rate convention applied to all other economies and re-pinned its five-sector value-added base (the China figure rises from $134 to $240 billion and the combined U.S.–EU–China total from $1.15 to $1.26 trillion accordingly). Signs and conclusions are unchanged throughout.

Disclaimer

This paper is an independent academic working paper prepared by the authors solely in their research capacity. It is submitted to SSRN for academic distribution and represents the authors' independent contribution to the economic literature on artificial intelligence and labor markets. This paper does not constitute investment advice, a trading recommendation, or a solicitation of any kind. Nothing in this paper should be construed as relating to any commodity interest account, agreement, or transaction, nor as an offer or invitation to invest in any fund, managed account, or trading program. It makes no reference to the past or projected performance of any trading program, commodity pool, or managed account, and contains no hypothetical performance results within the meaning of applicable commodity trading regulations. The views, projections, and conclusions expressed herein are solely those of the authors in their individual research capacity and do not represent the views or positions of Liminal Capital LLC, its principals, employees, affiliates, or clients. Institutional affiliations are disclosed for identification purposes only. The authors are employed by Liminal Capital LLC, a CFTC-registered Commodity Trading Advisor and NFA Member (NFA ID: 0525645). This relationship is disclosed in accordance with standard academic conflict of interest practice and does not imply that this paper was prepared in connection with Liminal Capital's commercial activities. All data sources used in the analysis are publicly available (Census Bureau BTOS AI Supplement, Eurostat `isoc_eb_ain2`, BLS OEWS and QCEW, O*NET-SOC, Felten-Raj-Seamans AIOE); no proprietary firm-level or trading data informs the empirical work. Forward-looking statements and projections in this paper are expressions of academic opinion subject to significant uncertainty and should not be relied upon for investment or financial decision-making. Past performance is not indicative of future results.

1. Introduction

There is substantial disagreement in the forecasting literature about the macroeconomic impact of artificial intelligence. Projections range from Acemoglu's (2024) 0.66 percent cumulative TFP gain over ten years to Hatzius et al.'s (2023) 1.5 percent annually, a spread of more than an order of magnitude. What these forecasts share is that they project impact from structural assumptions about diffusion rather than from measurement of what has already happened. Separately, a large post-pandemic literature has explained US productivity dynamics from 2020 through 2022 via compositional reallocation, labor-quality effects, and capital deepening, with little said about what has happened since the launch of ChatGPT in late 2022. This paper closes this gap by measuring from observed outcomes rather than projecting from structural assumptions.

Our measure of sectoral AI exposure is the Workflow Completeness construct (C*), introduced in Verschuere and Cameron (2026a) and validated against direct firm-reported adoption in Verschuere and Cameron (2026b). C* scores each occupation by the share of its work that AI is structurally capable of automating and performing autonomously, then aggregates to sectors using employment shares. Information and Professional Services score high; Construction and Industry score low. The validation work establishes that C* tracks observed AI adoption closely in both U.S. and European data, allowing us to use it as a proxy for AI deployment intensity. We apply C* to sectoral productivity data across nineteen economies (the United States, twelve EU economies, five non-EU OECD economies, and China) through 2025Q4.

This paper builds directly on Verschuere and Cameron (2026a), which established the U.S. productivity result using a cross-sectional γ estimator: a single post-period regression of sector-level excess productivity on C* across nine functional sectors. That estimator is statistically underpowered, as Verschuere and Cameron (2026a) note; with nine sector clusters it identifies the gradient but sits below what cluster-robust panel inference comfortably accommodates. The present paper provides the deeper methodological treatment: a panel difference-in-differences specification with sector and time fixed effects, a full inference cascade across cluster-robust, wild cluster bootstrap, and permutation methods, ten named alternative-mechanism tests,[4] and an extension to eighteen additional economies. The panel-DiD specification is more conservative in its identification than the cross-sectional γ estimator: by absorbing the common post-2022Q4 level shift into the time fixed effect, it

[4]The ten: pandemic-era continuation (Section 5.2); IT-investment payoff, cognitive task content, measurement artifact, over-hiring correction, and cost of capital (Section 5.3); date placebos, mean reversion, jobless-recovery dynamics, and the 2008 placebo (Section 7.1).

nets out the decay of pandemic-era productivity elevation rather than attributing it to AI, which the single-period cross-section cannot separate. The headline figure below is therefore a better-identified estimate of the same quantity that Verschuere and Cameron (2026a) anchored with the γ-based approach; the two are compared directly in Section 5.4, where the DiD-based and γ-based estimates are shown to bracket a common range.

We show that in the United States, AI is currently contributing approximately $878 billion per year to trend productivity growth at the end-of-2025 run-rate (95 percent CI [$602B, $1,155B]), raising the 5-sector aggregate productivity growth rate to +5.0 percent per year against a +2.2 percent counterfactual without AI. This figure is the end-of-2025 run-rate: the AI-attributable productivity flow at the rate prevailing in the final quarter of the sample, computed from the fitted DiD slope evaluated at end-of-period. It is computed on the five-sector primary panel (Information, Professional Services, Trade/Transport, Industry, Construction), which covers approximately 54 percent of U.S. nominal GDP at a value-added base of $16.0 trillion; we use this five-sector specification throughout the cross-country analysis and the dollar attribution because it is the cross-country-comparable subset on which Eurostat, OECD, and Chinese data align. The gains are concentrated in the sectors with the highest AI exposure (Information and Professional Services, with Finance also entering when the U.S. is examined on the fuller nine-sector panel) and accompanied by employment reductions in those same sectors, while low-exposure sectors show neither pattern. This output-up, employment-down signature is what AI-driven labor substitution looks like in aggregate sectoral data, and it is mechanically distinct from the pandemic-era productivity surge that preceded it, in which high-exposure sectors were the laggards on productivity and the leaders on employment growth. Ten alternative explanations for the post-2022 pattern are tested: pandemic-era continuation, delayed payoff to 2020–21 IT investment, tech-sector over-hiring correction, cost-of-capital exposure, date placebos, mean reversion, jobless-recovery dynamics, a 2008-recovery placebo, and — adopted from the companion papers — cognitive task content and measurement artifact (labor-quality and monetary-policy channels are tested within the pandemic-continuation and cost-of-capital analyses, Appendix C). None reproduces the joint signature; the closest competitor, tech-sector over-hiring, reduces the estimated effect by roughly a quarter when controlled for directly while leaving it significant (Section 5.3). The U.S. result is therefore not a candidate explanation among others but the joint pattern that the alternatives are inconsistent with.

The mechanism extends to other economies at smaller magnitudes that scale with deployment intensity. The twelve EU economies pooled show AI contributing approximately +1.55 percentage points per year to 5-sector productivity growth at the end-of-2025 run-rate, equiv-

alent to $138 billion per year (95 percent CI [$23B, $254B]; significant under analytic two-way clustering, borderline under the bootstrap; Section 6.2). China shows AI contributing +1.96 percentage points per year to 5-sector productivity at the end-of-2025 run-rate, equivalent to $240 billion (95 percent CI [$116B, $365B]), on top of an already-strong non-AI productivity growth path. (The U.S. figures quote realized versus counterfactual growth rates; the EU and China figures quote the AI contribution itself, so the percentages are not directly comparable across economies.) Across these three regions combined, AI is contributing approximately $1.26 trillion per year to trend productivity at the end-of-2025 run-rate, computed on the five-sector primary panel base of $37.1 trillion in value added (combined U.S. + EU 12 + China); aggregate dollar flows including the broader sector mix would be larger. A handful of additional individual economies (Germany and Sweden) show statistically significant positive contributions in their own right; Finland is marginal (one-sided WCB p = 0.093) and Australia reaches significance only when Agriculture is excluded, while the remaining major economies in the panel (Japan, the United Kingdom, Canada, South Korea) return positive but insignificant or weakly negative point estimates not statistically distinguishable from zero at current sample lengths. A striking feature of the cross-country picture is the role of the AI lift in containing what would otherwise be productivity stagnation: in six of twelve EU economies, as well as Japan and Australia, the counterfactual trend growth without AI is negative, with AI providing the majority of the post-ChatGPT productivity acceleration. Section 6 develops these findings in turn.

The findings carry implications across three policy domains. First, for monetary policy: an AI-driven productivity acceleration is a positive supply shock, calling for accommodation rather than restriction. The U.S. contribution at approximately 3.0 percent of nominal GDP per year is an order of magnitude larger than the supply shocks that have featured in Fed deliberations over the past decade; the equivalent in the EU 12 panel is materially smaller. Output-gap measures and Phillips-curve specifications that do not incorporate this expansion will systematically understate slack. Second, for fiscal policy: productivity-driven nominal GDP growth raises the (g - r) gap in the debt-sustainability identity without imposing the inflationary or contractionary costs of the alternatives. The largest fiscal-headroom benefit accrues to the United States, both in absolute and per-GDP terms; the EU 12 economies and the United Kingdom, where the productivity contribution is small or absent, will not realize comparable benefits from this channel. Third, for the measurement infrastructure on which AI policy is built: published digital-readiness composites, including the WIPO Global Innovation Index 2025 and the OECD DESI-equivalent composite, do not predict where AI is actually translating into measured productivity gains across the seventeen OECD economies for which both metrics are available. The methodology developed here, anchored in observed

sectoral productivity rather than theoretical readiness, provides a more reliable framework for ongoing monitoring of AI's economic effects across economies.

The paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the data. Section 4 develops the methodology. Section 5 presents the United States results: the headline finding and inference cascade, the joint pattern of output gains and employment reductions that distinguishes the post-ChatGPT regime from the pandemic-era one, the rejection of leading alternative mechanisms, and the dollar attribution. Section 6 presents the cross-country results: the principled sector restriction, the EU pooled panel, China, the non-EU OECD economies, the trend-growth decomposition with and without the AI contribution, and the failure of digital-readiness composites to predict observed outcomes. Section 7 reports robustness and extensions including falsifiers stated in advance of the data that can test them. Section 8 develops the policy implications. Section 9 concludes.[5]

2. Literature Review

The paper engages with three literatures: AI productivity forecasting, empirical measurement of AI productivity effects, and post-pandemic US productivity.

Forecasts of AI's productivity impact. Published forecasts of AI's near-term productivity contribution differ widely. Estimates range from Acemoglu (2024), who applies Hulten's theorem to task-level cost savings and projects a cumulative 0.66 percent increase in US TFP over ten years, to Hatzius et al. (2023), who project AI-driven productivity growth of 1.5 percent annually in the United States over the coming decade; intermediate estimates appear in Penn Wharton Budget Model (2023), McKinsey Global Institute (2023), and International Monetary Fund (2024). What these forecasts share, despite their numerical disagreement, is a forward-looking methodological structure: each projects impact from structural assumptions about diffusion and task substitutability. None tests whether the productivity signature its framework predicts is already present in aggregate data. The present paper takes the opposite approach, starting from observed outcomes.

Empirical measurement of AI productivity effects. A parallel empirical literature measures AI productivity effects in controlled settings and maps task-level AI capability to occupations. On the productivity side, Brynjolfsson, Li, and Raymond (2025) document a 15 percent increase in issues resolved per hour among customer-service agents deploying

[5]A companion methodology paper, Verschuere and Cameron (2026b), establishes the empirical equivalence between C* and direct firm-reported AI adoption underpinning the present paper's use of C* as a measurement proxy. The framework underlying C* construction is developed in Verschuere and Cameron (2026a), which also provides forward-looking GDP, labor-market, and inflation implications outside the scope of the present paper.

generative AI at a Fortune 500 firm; Noy and Zhang (2023) report a 40 percent time reduction and 18 percent quality gain in a writing-task RCT on mid-level professionals. On the task-exposure side, Frey and Osborne (2017) estimate that 47 percent of US employment faces high automation risk using expert scoring of O*NET tasks; Webb (2020) constructs an exposure measure from the overlap between patent text and job-task descriptions; Felten, Raj, and Seamans (2021) map AI-progress benchmarks to O*NET occupational abilities to produce the AIOE score; Eloundou et al. (2023) use GPT-4 to classify tasks by LLM substitution potential; Massenkoff and McCrory (2026) introduce an "observed exposure" measure that calibrates Eloundou et al. (2023) task-exposure scores against Anthropic Economic Index LLM usage logs and validate it against BLS occupational employment projections. The Massenkoff-McCrory construct is closely related to C^* in motivation but methodologically distinct: C^* derives task-level exposure from a theoretical workflow-completeness model validated against firm-reported adoption surveys, while Massenkoff-McCrory validates an existing theoretical exposure index against revealed LLM usage logs. We treat the two instruments as complementary; Verschuere and Cameron (2026b, Section 6.3) discuss the relationship in detail. These two strands jointly establish that AI delivers real productivity gains in specific deployments and that those gains concentrate in occupations with high task-level exposure. What neither strand establishes is whether these micro-level gains aggregate into measurable sector-level productivity changes in national accounts data. This paper fills this gap by providing the first cross-country sectoral measurement of AI's contribution to aggregate productivity growth both in the US and internationally.

Post-pandemic US productivity. A literature on post-COVID US productivity dynamics explains why US productivity rose above pre-2019 trend during 2020-2022. Gordon and Sayed (2022) attribute the surge to excess-layoff dynamics: firms cut hours sharply at the pandemic onset, producing a transitory productivity boom as the partial-and-lagged hours response unwound; Stewart (2022) documents that concentrated low-wage job losses produced labor-quality improvement via selective unemployment, a cleansing mechanism with foundations in Caballero and Hammour (1994); Fernald and Li (2022) and Fernald et al. (2024) attribute a substantial share to cyclical utilization, off-the-clock hours, and capital deepening during the pandemic; International Monetary Fund (2024) identifies a small AI-specific digital premium among teleworkable sectors. None of these mechanisms predicts a cross-sectional pattern that scales with task-level AI exposure. The post-2022Q4 period following the introduction of ChatGPT has accordingly received little attention in this literature as a regime requiring its own explanation.

3. Data

The analysis draws on sectoral productivity series for the United States and eighteen other economies, firm-reported AI adoption from the US Census Bureau and Eurostat, and supplementary measures for alternative-hypothesis tests.

Data sources. We construct sectoral productivity series from 2015 through 2025 from national accounts and labor market data. United States data combines BEA GDP-by-Industry real chained-dollar value added with FRED establishment employment by supersector, mapped to the NACE Rev 2 A10 classification via standard NAICS-to-ISIC crosswalks. EU data uses Eurostat `namq_10_a10` for chain-linked value added in 2015 reference-year volumes and `namq_10_a10_e` for employment on a persons basis. Non-EU OECD economies use national statistical agency series at quarterly frequency: ONS Quarterly National Accounts (United Kingdom), ESRI Quarterly Estimates (Japan), StatCan Industry Accounts (Canada), ABS Quarterly National Accounts (Australia), and BOK with OECD NAMAIN10 (South Korea). China uses annual data constructed from UN UNData Table 2.5 (constant-2015-prices real value added, 2015-2020), NBS Statistical Yearbook Table 3-6 (decomposition of the M_N, O-Q, R-T tertiary subgroups in the NBS grouping; R-T maps to A10 R–U), and NBS Q4 advance press releases (sectoral real growth rates, 2021-2025); Chinese sectoral employment is constructed from NBS Yearbook Table 4-6 sectoral shares applied to NBS three-strata totals.

Firm-reported AI adoption at the sector level comes from two sources. The US Census Bureau's Business Trends and Outlook Survey (BTOS) AI Supplement (Bonney et al. 2024) provides US data in two vintages: the 2023-24 vintage pools six biweekly waves from December 2023 through February 2024 with the narrow wording "In the last two weeks, did this business use Artificial Intelligence in producing goods or services?"; the 2025-26 vintage pools biweekly waves during 2025 and early 2026 with the broadened November 2025 wording "In the last two weeks, did this business use Artificial Intelligence in any of its business functions?" Eurostat `isoc_eb_ain2` provides 2024 sector-level AI adoption rates for 12 EU economies on five A10 sectors (J, M_N, G-I, B-E, F). Sector-level capital intensity for the pandemic-era IT investment horse race uses BEA Fixed Assets Table 3.7I; sector-level capital intensity for the cost-of-capital tests uses BEA Fixed Assets current-cost net stocks of private fixed assets by industry (Table 3.1ESI). Policy rates for the cross-country cost-of-capital tests are taken from central bank announcements and are cross-checkable against the BIS policy-rate statistics.

Country coverage. The cross-country panel includes 18 economies organized into three groups. Twelve EU economies have quarterly Eurostat A10 data through 2025Q4 (Germany,

France, Italy, Spain, Netherlands, Belgium, Austria, Sweden, Denmark, Finland, Norway, Portugal).[6] Five non-EU OECD economies have quarterly sectoral national accounts (United Kingdom, Japan, Canada, Australia, South Korea); South Korea uses OECD annual data with Boot–Feibes–Lisman interpolation (Boot, Feibes and Lisman 1967) for pre-2023 quarterly observations. China uses annual data 2015-2025. Ireland and Luxembourg are excluded because multinational profit-shifting (Ireland) and cross-border banking concentration (Luxembourg) produce sectoral series not comparable to other panel economies. Central European economies (Poland, Czechia, Hungary, Romania) are excluded because chain-linking and base-year revisions produce apparent sectoral discontinuities that reflect methodology rather than economics.

Sectoral classification. We use the NACE Rev 2 A10 classification throughout: Agriculture (A), Industry excluding construction (B-E), Construction (F), Trade, transport, accommodation and food services (G-I), Information and communication (J), Finance and insurance (K), Real estate (L), Professional, scientific and administrative services (M_N), Public administration, education and health (O-Q), and Arts, entertainment and other services (R-U). A10 is the most granular classification consistent across Eurostat, BEA, OECD, and the additional national statistical agencies used for non-EU economies. Five A10 sectors (J, M_N, G-I, B-E, F) are jointly observable in Eurostat AI adoption data and serve as the harmonized cross-country test scope (Section 6.1).

Replication scripts reproducing every table and figure in the paper are available from the authors and will be posted alongside the SSRN record.

4. Methodology

4.1 Workflow Completeness (C*)

We use the Workflow Completeness measure (C*) introduced in Verschuere and Cameron (2026a) as our exposure variable. C* is constructed from the occupational composition of each sector and the AI-completability of each occupation's tasks: each of 18,796 O*NET task statements is scored on a four-point scale reflecting the degree to which the task is AI-tractable, occupational C* is the importance-weighted mean normalized to [0,1], and sector-level C* is the employment-weighted average across occupations.[7] The companion

[6] We label this group the "EU 12" for brevity; it comprises eleven EU member states and Norway, an EEA member covered by the same harmonized Eurostat national accounts.

[7] Earlier vintages of this work used a smaller cross-section of scored occupations per sector; the current calibration deepens the construction to the full O*NET corpus under the top-3-occupation × top-10-task selection rule applied uniformly across sectors. The resulting sector-level C* values are moderately lower than those previously reported (most visibly in Information, Finance, and Professional Services, where the broader task pool dilutes the share of fully cognitive tasks). All headline regression results and identification

methodology paper (Verschuere and Cameron 2026b) validates C* against direct adoption measurement: U.S. BTOS firm-reported adoption correlates with C* at r = +0.89 across nine A10 sectors (near the r = +0.91 ceiling between two BTOS vintages); Eurostat-measured AI adoption in twelve EU economies correlates with C* at within-country R^2 of 0.58-0.76; and the methodologically independent Felten-Raj-Seamans AIOE measure converges with C* at r = +0.937, with predictive equivalence in panel specifications at 99 percent of the C* coefficient. We treat C* as a validated proxy for AI deployment intensity and refer to Verschuere and Cameron (2026b) for the validation evidence and to Verschuere and Cameron (2026a) for full construction details. For the cross-country analysis, C* is held constant across countries at the occupational level, the construct measures the ceiling of what AI is capable of doing to an occupation's tasks, which does not vary by country, while sectoral C* values differ across countries through country-specific occupational employment shares.

4.2 Excess Productivity

For each country c and sector s, we fit a log-linear pre-trend on 2015Q1-2019Q4 data:

$$\log \text{Prod}_{c,s,t} = \alpha_{c,s} + \beta_{c,s} \cdot t + \varepsilon_{c,s,t}, \quad t \in [2015\text{Q1}, 2020\text{Q1})$$

We then compute the post-period excess as the percentage deviation from the extrapolated trend:

$$\text{excess}_{c,s,t} = 100 \cdot \left[\exp\left(\log \text{Prod}_{c,s,t} - (\hat{\alpha}_{c,s} + \hat{\beta}_{c,s} \cdot t)\right) - 1\right]$$

This measure is unit-free, internationally comparable, and agnostic to the level of productivity in any given sector or country. Pre-period fit quality is strong, with mean R^2 of 0.73 across retained country-sector pairs. Sector-country pairs with negative pre-trend slopes are excluded because log-linear extrapolation is unreliable in that case; this affects a small number of country-sector combinations, primarily in Agriculture and Arts/Other services in a few European economies.

The post-period for excess measurement covers 2023Q1 through 2025Q4, twelve quarters following ChatGPT's public release in late November 2022; the DiD treatment indicator of Section 4.3 turns on at 2022Q4, the release quarter, giving thirteen treated quarters. We do not exclude the pandemic window (2020Q1-2022Q3) from the data; instead, we analyze

tests in this paper use these updated values; the sign and significance of every coefficient is unchanged, with only modest quantitative shifts in point estimates. The U.S. dollar attribution figures in Section 5.4 reflect these updated C* inputs.

pandemic-era dynamics explicitly as their own regime in Section 5.2 (joint pattern and regime separation) to establish that post-ChatGPT patterns are mechanistically distinct.

4.3 Panel Specification

If AI is driving post-ChatGPT productivity, two things should be visible in the data simultaneously. Productivity in high-C* sectors should grow faster than productivity in low-C* sectors after 2022Q4, and the gap should widen with time as adoption diffuses, not appear as a one-time level shift. We therefore need a specification that (i) compares sectors against each other cross-sectionally, (ii) compares the post-ChatGPT window against the pre-ChatGPT baseline, and (iii) separates the one-time level shift from the ongoing slope change. A two-way fixed effects difference-in-differences panel with both level and slope interactions in C* does all three.

The panel specification is:

$$\text{excess}_{s,t} = \alpha_s + \alpha_t + \Delta\alpha \cdot (C_s^* \cdot D_t^{\text{post}}) + \Delta\beta \cdot (C_s^* \cdot t \cdot D_t^{\text{post}}) + \varepsilon_{s,t}$$

where α_s absorbs sector-invariant heterogeneity, α_t absorbs common-to-all-sectors shocks each quarter (including the pandemic and post-pandemic aggregate productivity dynamics), t indexes quarters since 2015Q1, and $D_t^{\text{post}} = 1$ for $t \geq$ 2022Q4.

The two coefficients of interest are:

- $\Delta\alpha$: the immediate level shift at ChatGPT launch, scaled by C*
- $\Delta\beta$: the per-quarter rate at which the C*-scaled gap widens post-ChatGPT

Under the null of no differential AI effect, both coefficients are zero. The coefficient of primary interest is $\Delta\beta$, which captures the ongoing diffusion signature that logistic AI adoption would predict. $\Delta\alpha$ is a nuisance parameter that is mechanically correlated with $\Delta\beta$ (both share the factor $C^* \cdot D^{\text{post}}$); it is identified but not individually interpretable, and reported with full coefficients in Appendix B. We refer to this primary specification as Spec E throughout.

A fully-interacted robustness specification, which absorbs sector-specific pre- and post-period level shifts into separate parameters, is reported in Appendix A. Spec A returns a larger headline coefficient (+3.381 versus the primary +2.09) because it isolates the C*-scaled slope from the level-shift component that the primary specification absorbs into a single $\Delta\alpha$ parameter; the two specifications reach the same qualitative verdict on the US result and produce directionally consistent between-sector orderings when evaluated at the observed C* range. The primary specification exhibits greater robustness to wild cluster bootstrap inference because of its parsimony (the fully-interacted model's additional parameters degrade small-cluster

bootstrap reliability) without losing any of this economic content.

Panel size per country is 396 sector-quarter observations (9 sectors × 44 quarters; Real estate (L) is excluded because imputed rents dominate its measured value added, while Agriculture is retained, with robustness to its exclusion reported in Section 7.2).

4.4 Inference

We report four variants of standard errors for the primary specification. Each is increasingly conservative with respect to the small-cluster dimension inherent to the 9-sector panel (G = 9 sector clusters after dropping Real estate): OLS homoskedastic, sector-clustered, time-clustered, and Cameron-Gelbach-Miller (2011) two-way clustered.

Cluster-robust inference with G = 9 cross-sectional clusters is near the lower limit of reliability (Cameron and Miller 2015; MacKinnon and Webb 2017). We therefore supplement asymptotic cluster-robust standard errors with wild cluster bootstrap inference at B = 9,999 draws under both Rademacher and two-point Mammen weight distributions. Within the bootstrap family, Rademacher weights with sector clustering are the designated primary configuration: Rademacher is the standard default weight choice in applied practice, and the sector dimension is the conservative clustering dimension for this panel; the alternative weight and clustering variants are reported alongside as sensitivity. We additionally implement a permutation test on C* at B = 10,000 permutations that shuffles C* values across sectors and re-estimates under the sharp null that C* has no differential predictive power. The permutation test is our most demanding significance statement because it requires only exchangeability of C* labels, with no cluster-asymptotic or bootstrap-weight assumption. Implementation details are in Appendix E.

Which inference method for which task. The four methods above answer different statistical questions and are appropriate for different purposes. We use them deliberately throughout the paper as follows. *For hypothesis testing of slope coefficients under the sharp null H_0: $\Delta\beta = 0$*, we report wild cluster bootstrap and permutation p-values as the primary inference, because both are valid under small G and require minimal distributional assumptions. The sector-clustered analytic standard error is reported alongside for transparency but should be interpreted as approximate at G = 9. *For interval estimation of slope coefficients and for propagating uncertainty through the dollar-attribution formula*, we use the analytic sector-clustered standard error, not WCB. The reason is that WCB is constructed to produce valid *tests* (p-values) under small G, but the simulation-based interval estimates it produces are pessimistic for parameters not on the null: the WCB CI is constructed by inverting tests across a grid of null values, and the small-cluster correction that makes WCB conservative for

hypothesis testing also makes WCB CIs systematically wider than the asymptotically appropriate interval. For point estimation of the dollar contribution at end-of-2025, the analytic sector-clustered CI is the correct interval-estimation tool; we note where the choice matters substantively. *For placebo tests and date placebos in Appendix C*, we report both inference methods the placebo module computes (OLS and two-way clustered) rather than a single preferred method, because placebos are diagnostic tests where consistency across methods is itself informative: a placebo that is insignificant under WCB and two-way clustering but borderline under sector clustering is genuinely ambiguous evidence, and we present it that way.

4.5 Identification Strategy: The Joint-Pattern Argument

Our causal identification rests on a specific prediction about what pattern AI adoption should produce in sectoral national accounts data, and tests whether that pattern is present.

If AI adoption is driving productivity changes, four features should appear jointly: (i) positive cross-sectional scaling of post-adoption productivity with C*; (ii) negative cross-sectional scaling of post-adoption employment with C*, reflecting labor substitution; (iii) emergence at 2022Q4 with broad-based generative AI deployment, not at arbitrary earlier or later dates; (iv) a cross-sectional fingerprint mechanistically distinct from pandemic-era productivity dynamics, which the literature attributes to compositional reallocation, labor quality effects, and capital deepening, none of which scale with task-level AI exposure.

The identification argument is therefore not "correlation is consistent with causation." It is: a specific multi-feature joint pattern consistent with AI adoption is present in US national accounts data, is not explained by any named alternative, is measured using an instrument validated against direct adoption data, and appears in cross-country data with magnitudes that scale with the intensity of AI deployment, strongest in the US, second-largest in China, smaller but statistically significant in the EU pooled panel, and absent or weaker in economies with smaller measured deployment.

A note on $\Delta\beta$ units. The slope coefficient $\Delta\beta$ is reported throughout this paper in units of *percentage points per quarter per unit of C**, since the panel is quarterly and the dependent variable is in percentage points. When we annualize for headline comparability or trend-growth interpretation, we multiply by 4 (e.g., $\Delta\beta = +2.088$/qtr corresponds to $+8.35$/yr). Cross-country tables that report annualized values explicitly note "annualized" in the column header. Dollar attributions multiply $\Delta\beta \times$ annualization $\times$ C*-weighted-VA, giving an annualized flow.

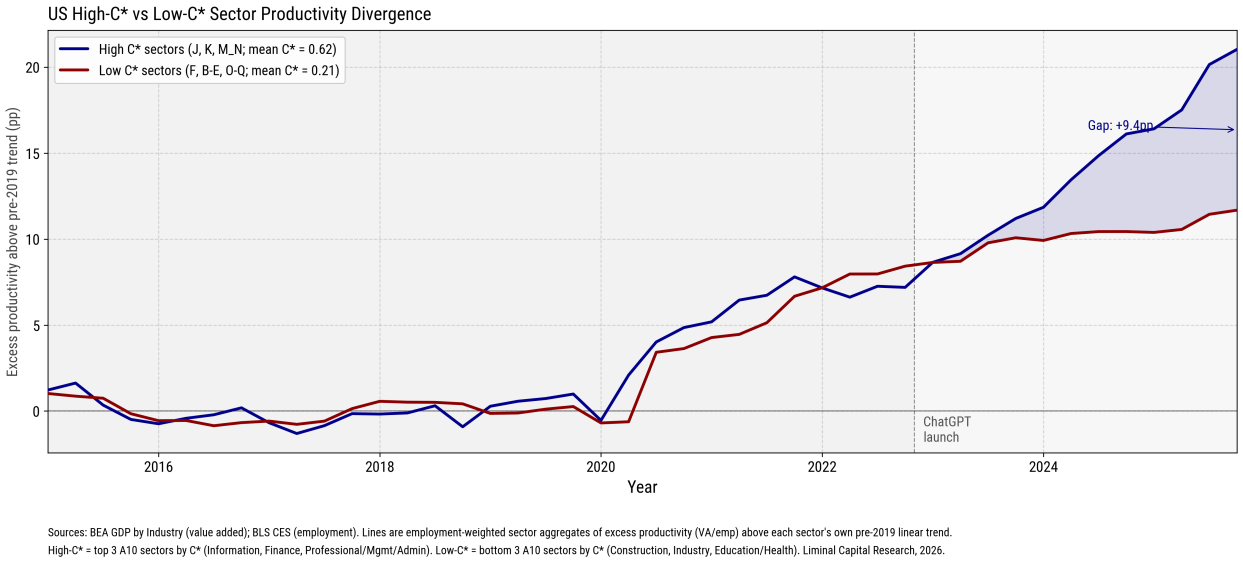


Figure 1: High-C^* vs Low-C^* Sector Productivity Divergence, US 2015Q1–2025Q4. "High-C^*" = the three A10 sectors with the highest C^* (J Information, K Finance, M_N Professional/Management/Administrative; mean $C^* = 0.62$); "Low-C^*" = three low-C^* A10 sectors (F Construction, B-E Industry, O-Q Education/Health; mean $C^* = 0.21$; G-I, third-lowest by C^*, is excluded from this illustrative grouping — the panel regressions use all sectors). Each line is an employment-weighted aggregate of sector-level excess productivity (VA/emp) above each sector's own pre-2019 linear trend. Vertical dashed line marks the 2022Q4 ChatGPT release.

5. US Results

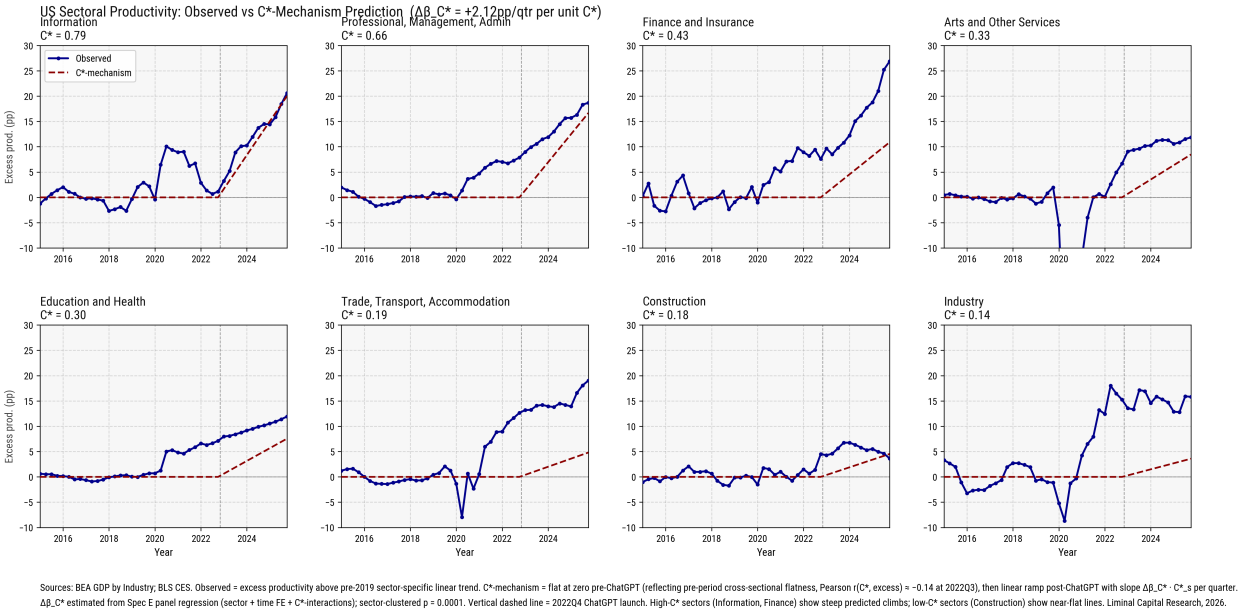
This section reports the United States results: the panel regression coefficient and inference cascade (5.1), the joint signature of output gains coinciding with employment reductions in the same sectors (5.2), the rejection of named alternative mechanisms (5.3), and the dollar attribution of the productivity acceleration (5.4). Cross-country results are reported in Section 6.

5.1 Panel Coefficient and Inference Cascade

We test whether the US sectoral productivity pattern exhibits the joint signature consistent with AI adoption. Recall the panel specification from Section 4.3:

$$\text{excess}_{s,t} = \alpha_s + \alpha_t + \Delta\alpha \cdot (C_s^* \cdot D_t^{\text{post}}) + \Delta\beta \cdot (C_s^* \cdot t \cdot D_t^{\text{post}}) + \varepsilon_{s,t}$$

where $\Delta\beta$, the coefficient of primary interest, measures the per-quarter rate at which the C*-scaled productivity gap widens after 2022Q4. The panel estimate is $\Delta\beta = +2.088$, statistically significant under the analytic cluster-robust methods (sector-clustered, time-clustered, two-way), the primary wild cluster bootstrap, and permutation inference; OLS,



Sources: BEA GDP by Industry; BLS CES. Observed = excess productivity above pre-2019 sector-specific linear trend. C*-mechanism = flat at zero pre-ChatGPT (reflecting pre-period cross-sectional flatness, Pearson r(C*, excess) = −0.14 at 2022Q3), then linear ramp post-ChatGPT with slope Δβ_C* · C*_s per quarter. Δβ_C* estimated from Spec E panel regression (sector + time FE + C*-interactions); sector-clustered p = 0.0001. Vertical dashed line = 2022Q4 ChatGPT launch. High-C* sectors (Information, Finance) show steep predicted climbs; low-C* sectors (Construction) show near-flat lines. Liminal Capital Research, 2026.

Figure 2: Observed vs Fitted Excess Productivity by Sector under Spec E, 2015Q1–2025Q4 (eight non-Agriculture sectors; the in-figure slope, +2.12, is the excluding-Agriculture estimate of Table 20)

which does not adjust for the panel structure, sits at p = 0.09 and the most conservative bootstrap variant (Mammen weights, sector clusters) at p = 0.11 (Table 3).

Table 1: US panel $\Delta\beta$ under the primary specification

Coefficient	Estimate	Interpretation
$\Delta\alpha$	nuisance	Level shift at 2022Q4 scaled by C* (reported in Appendix B)
$\Delta\beta$	**+2.088**	Per-quarter rate of C*-scaled gap widening post-ChatGPT

The point estimate on $\Delta\beta$, the coefficient of primary interest, is +2.088 percentage points per quarter per unit of C*. Evaluated across the Construction–Information contrast (C* from 0.178 to 0.787), this implies that Information productivity grows approximately 1.27 percentage points per quarter faster than Construction productivity due to the C*-scaled component alone, over and above any common post-ChatGPT shock absorbed by the time fixed effect.

A positive $\Delta\beta$ establishes that post-ChatGPT productivity scales with C* at a rate not observed pre-ChatGPT. We turn now to inferential robustness, then in Section 5.2 to the joint cross-sectional pattern that distinguishes the post-ChatGPT regime from the pandemic-era one, in Section 5.3 to the rejection of the leading alternative mechanisms, and in Section 5.4 to the aggregate dollar magnitude.

Cluster-robust inference with G = 9 cross-sectional units is near the lower limit of what

asymptotic estimators handle reliably. We therefore report the full inference cascade described in Section 4.4. A consolidated summary of every headline result under every inference method is provided in Table 18 (Appendix B).

Table 2: US $\Delta\beta = +2.088$ under four inference methods

Method	p-value	Notes
OLS homoskedastic	0.090	Not primary; ignores panel structure
Sector-clustered	< 0.001	Allows arbitrary within-sector autocorrelation
Time-clustered	< 0.001	Allows arbitrary within-quarter cross-sectional dependence
Two-way clustered	0.005	Cameron-Gelbach-Miller (2011) combined estimator

All three cluster-robust methods clear conventional significance thresholds. OLS without cluster adjustment falls short of conventional significance (p = 0.09). The divergence between OLS and clustered standard errors is not an efficiency property of the estimator — the point estimate is identical under every method — but a property of the variance formula: the clustered variance adds the within-cluster covariances of residual-score products, and the sign of those covariances determines the direction of the adjustment. For the headline regressor, whose within-sector variation is a smooth deterministic ramp, the covariances are on net negative and the clustered SE falls below the OLS SE; for the date placebos in Appendix C.3 they are positive and clustering roughly doubles the OLS SEs. Both directions are mechanical consequences of the same formula at G = 9, which is why headline testing leans on wild cluster bootstrap and permutation inference rather than on any analytic clustered variance.

Supplementing cluster-robust asymptotic inference with wild cluster bootstrap at B = 9,999 draws, under both Rademacher and two-point Mammen weight distributions:

Table 3: US $\Delta\beta = +2.088$ under wild cluster bootstrap (B = 9,999)

Weight distribution	Cluster dimension	p-value	Monte Carlo 95% CI on p
Rademacher	Sector	0.016	[0.013, 0.018]
Rademacher	Time	0.053	[0.049, 0.057]
Mammen (two-point)	Sector	0.111	[0.105, 0.117]
Mammen (two-point)	Time	0.036	[0.032, 0.040]

Rademacher weights with sector clustering are designated the primary configuration on methodological grounds: Rademacher is the standard default weight for the wild cluster bootstrap, sector is the economically conservative clustering dimension under the inference framework of Section 4.4, and at $G = 9$ the Rademacher distribution provides $2^9 = 512$ distinct sign patterns, sufficient for inference at conventional levels; the two-point Mammen variant is reported as the conservative alternative. The designation follows the methodology, not the realized p-values.

The primary WCB configuration (Rademacher weights, sector clusters) returns p = 0.016, consistent with the analytic sector-clustered inference, and Mammen time-cluster inference agrees (p = 0.036). The two remaining variants are weaker: Rademacher time-cluster sits just above the 5 percent boundary (p = 0.053) and Mammen sector-cluster is the most conservative at p = 0.111, reflecting the sensitivity of bootstrap weight choice with G = 9 clusters (MacKinnon and Webb 2017). Headline inference therefore rests on the Rademacher sector-cluster configuration — designated primary on the methodological grounds of Section 4.4 rather than by prior registration — together with the permutation test below; all four variants are reported for transparency.

The most demanding inferential test is permutation inference on C*, which makes no cluster-asymptotic or bootstrap-weight assumption and tests the sharp null that C* has no differential predictive power across sectors.

Table 4: US $\Delta\beta = +2.088$ under C* permutation inference (B = 10,000)

Specification	Observed t-statistic	One-sided permutation p	Rank of observed t
Spec E (primary)	+1.70	0.0056	56 of 10,001
Spec A (robustness, Appendix A)	+3.58	0.0013	13 of 10,001

The observed US t-statistic ranks 56th of the 10,001-value null distribution from 10,000 random C*-label reassignments under the primary specification. The permutation p-value is 0.0056, of the same order as the WCB p-values, consistent with the expected direction of the relationship between permutation and WCB when cluster counts are small. WCB with G = 9 has only 2^9 = 512 distinct sign-flip patterns, producing a coarse discrete null distribution; permutation with 9! = 362,880 possible label assignments produces a finer null distribution and sharper inference.

The result is that $\Delta\beta = +2.088$ in US data clears the primary inference configuration (Rademacher/sector wild cluster bootstrap, p = 0.016, and permutation, p = 0.0056) and every analytic cluster-robust method (sector-clustered p = 0.0002, time-clustered and two-way below 0.01). The most conservative bootstrap variant (Mammen weights on sector clusters) sits at p = 0.111 and unadjusted OLS at p = 0.090 (Table 2); as Section 4.4 discusses, we treat the Rademacher/sector configuration and the permutation test as primary and report the full battery in Table 3 for transparency.

5.2 The Joint Pattern: Regime Separation from Pandemic-Era Productivity

The most direct alternative interpretation of the post-2022Q4 pattern is that it continues the pandemic-era productivity dynamics that existing literature attributes to compositional re-

allocation, labor-quality cleansing, and capital deepening. A sectoral decomposition of value added and employment across the two windows shows that the cross-sectional fingerprint reverses sign on both variables: post-ChatGPT productivity dynamics are mechanistically distinct from the pandemic regime that preceded them.

Table 5: Pandemic-window (2019Q4 → 2022Q3) sectoral decomposition, excl. Agriculture

Sector	C*	ΔVA (%)	ΔEmp (%)	ΔProductivity (%)
B-E (Industry)	0.141	+25.2	−0.2	**+25.4**
G-I (Trade/Transport)	0.189	+24.5	+2.3	+21.8
M_N (Professional services)	0.655	+23.1	+5.7	+16.5
K (Finance)	0.428	+17.8	+3.1	+14.2
R-U (Arts/Other)	0.333	+13.5	+0.5	+13.0
O-Q (Education/Health)	0.297	+11.4	−0.7	+12.1
F (Construction)	0.178	+15.0	+3.6	+11.0
J (Information)	0.787	+18.8	+7.4	+10.6
Correlation with C*		+0.02	**+0.79**	**−0.46**

During the pandemic window, productivity gains were not ranked by C* (r(ΔProductivity, C*) = −0.46) and high-C* sectors were *adding* employment (r(ΔEmp, C*) = +0.79). Information, the highest-C* sector, showed the weakest productivity growth. The pandemic-era gains concentrated in Industry, Trade/Transport, and Education/Health: sectors where excess-layoff dynamics (Gordon and Sayed 2022), labor-quality cleansing (Stewart 2022), and capital deepening (Fernald and Li 2022; Fernald et al. 2024) dominated, none of which scales with task-level AI exposure.

In the post-ChatGPT window (2022Q3 → 2025Q4) both correlations reverse: r(ΔProductivity, C*) = +0.86, r(ΔEmp, C*) = −0.60. Information becomes the leading productivity gainer at +37.3 percent, through +26.3 percent VA growth combined with −8.0 percent employment change. The mechanical signature across high-C* sectors is output-up, labor-down, productivity-up, sharply scaled with C*. This signature is consistent with independent labor-market evidence: Levanon (2026) documents an aggregate employment shortfall of approximately three million jobs concentrated in high-AI-exposure white-collar sectors (his FI-IPB aggregate of Finance, Insurance, Information, and Professional and Business Services), arising from continued growth in sectoral output alongside declining or stagnant headcount since late 2022.

The quarter-by-quarter cross-sectional regression confirms the regime change. Through 2022Q3, the origin-forced slope coefficient γ (excess productivity per unit of C*, the paper-scale metric[8]) ranges from −1.44 to +5.24 with mixed significance; under standard Pearson

[8]Two γ scale conventions appear in this paper. The *A10 scale* measures γ directly on the cumulative

correlation the relationship is noisy in the pre-period (r ranges from -0.49 to $+0.53$) because of high cross-sectional dispersion in raw productivity deviations. Post-ChatGPT the γ slope builds monotonically, reaching $\gamma = +14.35$ ($t = +5.24$, $p = 0.001$) by 2025Q4 with a clean C*-gradient across the eight non-Agriculture sectors.

The regime-separation evidence directly rejects the dominant macroeconomic interpretations of the pandemic-era productivity surge as candidates for the post-ChatGPT phenomenon: the cross-sectional fingerprint reverses sign on both productivity-C* and employment-C* correlations, a pattern incompatible with continuation of mechanisms that do not scale with task-level AI exposure. The pandemic-era accounts remain valid for 2020-2022; the post-2022Q4 phenomenon operates through a different mechanism. Agriculture is excluded from the decomposition and cross-sectional correlations above because its pandemic-era readings track the 2020-2022 agricultural commodity price cycle rather than productivity dynamics; the post-ChatGPT panel result is robust to its inclusion, and the full argument is presented in Section 7.1.

COVID-absorbed robustness. As a direct identification check, we re-estimate the post-ChatGPT slope β_P in a specification that explicitly absorbs an arbitrary COVID-period level shift (a 2020Q1–2022Q3 indicator interacted with C^* is added to Spec E, and β_P denotes the post-treatment slope on $C_s^* \cdot t_P \cdot D_t^{\text{post}}$ with the clock t_P starting at treatment), removing any possibility that the slope coefficient is contaminated by uncorrected pandemic-era elevation. The headline finding survives: pooled US $\beta_P = +8.21$ pp/yr per unit of C^* ($p < 0.001$); EU 12 pooled $\beta_P = +3.39$ pp/yr ($p = 0.0004$); all 18-economy pooled (the eighteen quarterly panels; China's annual panel excluded) $\beta_P = +2.84$ pp/yr ($p < 0.001$). The post-ChatGPT slope acceleration is not a COVID-level-shift artifact.

Dependence of pandemic-window magnitude on Agriculture. The pandemic-era panel slope $\Delta\beta_C^* = -8.01$ (estimated in the Section 5.3 horse-race specification that includes the IT-intensity control; the C*-only pandemic slope is -7.69) is partly driven by Agriculture's anomalous 2020-2022 productivity spike (+94 percentage points of excess at 2022Q3, attributable to the agricultural commodity price cycle rather than to AI-related dynamics).

excess-productivity basis of Section 4.2, in percentage points of cumulative deviation from pre-2019 trend per unit C^*. The *paper scale* rescales to the productivity-index basis of Verschuere and Cameron (2026a) for comparability with that paper's γ estimates; the empirical conversion in this sample is paper = A10 $\times\, 8.32/20.58 \approx 0.404$, the ratio of the cross-sectional gradient measured on the Verschuere and Cameron (2026a) excess-growth basis ($\gamma = 8.32$) to the same calibration cross-section's gradient measured on this paper's cumulative-excess basis ($\gamma = 20.58$); the 2025Q4 level gradient of +35.5 A10-scale thus equals +14.35 paper-scale. Unless tagged otherwise, γ values in the main text are paper-scale; Figure 4 and Appendix C.5 report A10-scale values. Origin-forced level gradients, origin-forced increment gradients, and event-study γ_t deviations (referenced to 2019Q4) are distinct objects and are labeled where reported.

Excluding Agriculture reduces the pandemic-window slope from −8.01 to −3.54 (a 56 percent attenuation, p = 0.009 sector-clustered; see Section 7.2). Two implications: (i) the post-ChatGPT slope $\Delta\beta = +2.088$ is essentially unchanged when Agriculture is excluded (becoming +2.12), so the headline does not depend on this sector; (ii) the cross-sectional sign-reversal documented in Table 5 is computed on the 8-sector panel that excludes Agriculture from the start, so it is not an Agriculture artifact. The 8-sector pandemic slope of −3.54 (without Agriculture) is still opposite in sign to the post-ChatGPT slope of +2.12 (without Agriculture), and the joint-pattern argument therefore holds in either inclusion convention. We present the Agriculture-included headline magnitude transparently because the alternative pandemic-era accounts (Gordon-Sayed, Fernald-Li, labor-quality cleansing) make claims about the full economy, not the non-agricultural subset.

Three-phase decomposition of the high-C* productivity trajectory. The event-study representation of the panel (Figure 5 in Appendix B) decomposes the post-2019 trajectory of high-C* productivity into three economically distinct phases, each with an independent labor-market mechanism. **Phase 1, Stimulus boom (2020Q1–2021Q4):** Information VA rose +16.4 percent, employment +2.5 percent, productivity +13.6 percent. Professional services moved similarly (VA +16.5, employment +2.5, productivity +13.7). Demand for digital services from remote work and stimulus disbursement scaled output well ahead of headcount. **Phase 2, Over-hiring hangover (2021Q4–2022Q4):** Information VA growth slowed to +3.8 percent while employment growth accelerated to +5.0 percent; productivity fell by −1.1 percent. This is the textbook over-hiring signature: in the year preceding the major tech layoffs of 2022–2024, J-sector firms grew headcount faster than they grew output. Professional services showed a milder version of the same pattern (VA +7.5, employment +3.5, productivity +3.9). **Phase 3, Post-ChatGPT (2022Q4–2025Q4):** Information VA grew +24.2 percent while employment fell −8.2 percent, producing a +35.3 percent productivity gain. Professional services: VA +18.7, employment −1.8, productivity +20.9. The 3-year sequence of Phase 3 corresponds to the well-documented round of tech-sector layoffs (announced by Meta, Google, Amazon, Salesforce, and others between 2022Q4 and 2024Q3) combined with AI deployment driving output. Three economically independent mechanisms, three sequential phases, each with verifiable labor-market evidence: this is the structural story underlying the cross-sectional fingerprint of Table 5 and the panel slope coefficient of Section 5.1.

5.3 Alternative Mechanisms

Beyond pandemic-era continuation, the post-2022 sectoral signature requires direct engagement with several further candidate mechanisms, treated in turn in this subsection: pandemic-

era IT investment payoff, cognitive task content, instrument-construction artifact, over-hiring correction, and the cost of capital. Four additional alternatives (date placebos, mean reversion from pandemic elevation, jobless-recovery dynamics, and post-recession comparison via a 2008 placebo) are tested and reported in Section 7.1, and the cost-of-capital alternative is treated at the end of this subsection with full results in Appendix C.8; none produces the observed joint pattern.

Pandemic-era IT investment. A serious empirical alternative is that C* correlates with 2020-2021 IT investment intensity, with post-2022 productivity gains reflecting the payoff to pandemic-era digital capital accumulation rather than AI adoption specifically. C* is high for Information, Finance, and Professional Services, and these same sectors invested heavily in IT during the 2020-2021 remote-work pivot.

We construct a sector-level IT-intensity measure from BEA Fixed Assets Table 3.7I (Investment in Private Intellectual Property Products by Industry). For each A10 sector:

$$\text{IT}_s = \frac{\frac{1}{2}(\text{IPP investment}_{s,2020} + \text{IPP investment}_{s,2021})}{\frac{1}{2}(\text{VA}_{s,2020} + \text{VA}_{s,2021})}$$

mapped from the BEA 96-industry hierarchy to A10 via value-added weighting and then standardized to a z-score IT^z across the ten sectors.

The Pearson correlation between C* and IT^z is +0.687 (Spearman +0.617), high enough that the two measures share substantial variation, but low enough that a horse race has meaningful discriminatory power. Manufacturing (B-E) has IT^z = +1.00 against the panel-minimum C* = 0.14 because R&D investment inflates IPP; Real Estate (L) has IT^z = −0.80 against C* = 0.34 because imputed rents inflate the VA denominator. These sectors provide the independent variation that allows the horse race to separate C* from IT^z.

The horse race extends the primary specification with parallel level and slope interactions for IT^z:

$$\text{excess}_{s,t} = \alpha_s + \alpha_t + \Delta\alpha_{C^*}(C^*_s D^{\text{post}}_t) + \Delta\beta_{C^*}(C^*_s t D^{\text{post}}_t) + \Delta\alpha_{\text{IT}}(\text{IT}^z_s D^{\text{post}}_t) + \Delta\beta_{\text{IT}}(\text{IT}^z_s t D^{\text{post}}_t) + \varepsilon$$

The central finding is that $\Delta\beta$_C* is essentially unchanged when IT intensity is included as a control; it shifts negligibly from +2.088 to +2.051. $\Delta\beta$_IT is approximately zero and insignificant (+0.013, sector-clustered p = 0.90). Under permutation inference the observed $\Delta\beta$_C* = +2.051 corresponds to a t-statistic rank of 264 out of 10,000 C* permutations,

Table 6: $\Delta\beta$_C* in horse race with IT intensity control

Specification	$\Delta\beta$_C*	$\Delta\beta$_IT	Sector-clust p ($\Delta\beta$_C*)	Permutation p ($\Delta\beta$_C*)
Baseline (C* only)	+2.088	—	< 0.001	0.0056
Horse race (C* + IT^z)	**+2.051**	+0.013	**0.0003**	**0.026**

giving one-sided p = 0.026. The pandemic-era IT investment alternative is not consistent with the data: IT intensity does not explain the post-ChatGPT C*-scaled pattern, and the narrower interpretation, that productivity gains are the delayed payoff of 2020-2021 digital investment rather than contemporaneous AI adoption, is not supported by the horse-race specification.

Cognitive task content (OVB). A distinct concern is that C^* is mechanically correlated with cognitive task content, which is in turn correlated with sectoral wage levels, raising the possibility that the post-ChatGPT signal reflects generic knowledge-intensive-sector dynamics rather than AI-specific exposure. Verschuere and Cameron (2026a, Section 4.5) provide the cleanest test of this through a discriminant-validity horse race at NAICS-2 granularity ($n = 19$), where C^* and log sectoral mean wage are highly collinear ($r = +0.875$) but the sample is large enough to identify the AI-specific component: in the joint regression C^* retains its full explanatory power ($\beta = +1.11$, $p = 0.002$) while log mean wage carries none ($\beta = -0.35$, $p = 0.25$). The C^* coefficient actually *grows* when wage is included as a control (from +0.808 univariate to +1.111 in the joint regression), the opposite of what wage-proxies-cognitive-intensity would predict. We do not duplicate this test in the present paper's panel-productivity specification because at the panel-relevant aggregation (A10, $n = 8$ excluding Agriculture), C^* and log wage are collinear at $r = +0.77$ but the sample provides too few degrees of freedom to separate two near-collinear regressors in the slope dimension, exactly as Verschuere and Cameron (2026a) document in their A10 secondary check. The NAICS-2 discriminant-validity test in that paper is the relevant evidence that the panel signal is AI-specific rather than cognitive-content generic.

Measurement artifact. The companion methodology paper (Verschuere and Cameron 2026b) addresses this directly: substituting direct BTOS firm-reported adoption for C* produces equivalent panel regression coefficients within sampling variability, and the methodologically independent Felten-Raj-Seamans AIOE measure produces a standardized effect 99 percent identical to the C* baseline. The post-2022 productivity signature is not driven by anything specific to the C* construction.

Tech-employment correction / over-hiring. A widely cited alternative reading of the post-2022Q4 productivity signature is that it reflects a mechanical correction of 2020-2022

over-hiring in high-C^* sectors (the late-2022/early-2023 tech layoff wave at Meta, Amazon, Google, Salesforce, and others), with VA/employment rising as the denominator contracts. Verschuere and Cameron (2026a) Section 4.2 tests this directly using sectoral employment levels rather than the panel DiD: high-C^* sectors averaged +5.15 percent above their pre-pandemic baseline during 2020-2022, low-C^* sectors averaged +5.0 percent, a difference of 0.15 percentage points ($p = 0.97$). The two groups hired identically before the post-ChatGPT divergence; the groups that started from statistically identical pre-divergence positions diverged by 6.6 percentage points through 2025Q3 (the companion paper's sample cutoff). The cleanest discriminant is the cross-sectional inversion test: Transport and Logistics ($C^* = 0.19$) had the largest hiring surplus in the entire dataset, +15.7 percent above its pre-pandemic peak in 2022, more than twice Information and Technology's +7.4 percent. Under the over-hiring-correction story, Transport should have produced the largest post-2022Q4 productivity correction; instead, it added $\sim$ 25,000 jobs post-ChatGPT and produced excess productivity of only +3.2pp, while Information cut headcount by 8 percent and produced +10.9pp of excess productivity. The sectoral pattern is opposite to what an over-hiring-correction story predicts. The same prior paper rules out the related "pandemic IT-investment payoff" reading via the IT-intensity horse race summarized in the previous paragraph. We adopt these findings here rather than redo the sector-level employment tests on a panel-DiD framework that is less sensitive to absolute headcount patterns; the panel DiD is appropriate for identifying the C^*-graded productivity slope but is not the right specification for testing whether the absolute employment paths distinguish AI from over-hiring correction.

Two further tests engage the over-hiring story within the panel framework itself, on the five-sector primary panel. First, a split-window test: a bounded one-time stock correction predicts a strong slope during the layoff wave and a weak or zero slope after it concludes, whereas AI diffusion predicts persistence. The post-ChatGPT slope is +2.00 per quarter over the layoff window (2022Q4-2024Q1, $p = 0.007$) and +1.79 over the post-layoff window (2024Q2-2025Q4, $p = 0.007$), a late-to-early ratio of 0.89: the slope persists essentially undiminished after the layoff wave ends. Second, a direct horse race that adds level and slope interactions in sector-level employment overshoot (the 2022Q3 deviation of sector employment from its 2015-2019 log-linear trend) to the primary specification. The two regressors are highly collinear in the cross-section ($r = +0.71$), so the horse race has limited resolving power, and we report it transparently: the overshoot slope enters positively (+5.13, $p = 0.001$) and the C^*-scaled slope attenuates by 28 percent to +1.42 while remaining significant at $p < 0.0001$. Read jointly with the inversion test above (the largest-overshoot sector, Transport, produced no post-2022Q4 correction while modest-overshoot Information

produced the largest productivity gain) and the split-window persistence, the attenuation is consistent with the collinearity of the two measures rather than with a bounded over-hiring correction, which the timing and cross-sectional evidence both contradict.

Cost of capital. The most common alternative reading of the post-2022 timing is monetary rather than technological: the 2022-2023 tightening raised the cost of capital, forcing labor shedding and capital discipline that mechanically raise value added per worker. The mechanism fails on each of four margins: three reported in Appendix C.8 and the sector rate-sensitivity horse race of Appendix C.4. Cross-country, the tightening was nearly synchronized (cumulative hikes of +300 to +525 basis points across the panel's hiking economies, standard deviation 48 bp), and no dose-response exists in either the change or the level of policy rates (six specifications, all p $\geq$ 0.80); the non-hikers bracket the effect distribution, with Japan among the weakest effects and easing China the second strongest. Within the United States, horse races against sector rate sensitivity (Appendix C.4) and against capital intensity under two definitions (C.8) leave the C*-scaled slope intact ($\hat{\Delta\beta}_K = +0.01$, p = 0.87, against $\hat{\Delta\beta}_{C^*} = +2.10$, permutation p = 0.015). And the timing runs the wrong way: between the 2023Q4 policy peak and 2025Q4 the federal funds rate fell 1.4 percentage points while the C* gradient rose from +8.6 to +14.35 (paper scale), the opposite of what a cost-of-capital channel predicts.

5.4 Dollar Attribution

This subsection translates the U.S. panel coefficient into an annualized dollar magnitude. The headline figure is **$878 billion per year (95% CI [$602B, $1,155B])**, the AI-attributable U.S. productivity flow at the end-of-2025 run-rate under the DiD-based trend-growth decomposition, computed on the five-sector primary panel (J Information, M_N Professional, G-I Trade/Transport, B-E Industry, F Construction) covering approximately 54 percent of U.S. nominal GDP. The run-rate is the AI-attributable flow at the rate prevailing in the final quarter of the sample; because the effect ramps roughly linearly from near zero at ChatGPT's release to its full strength by end-2025, the run-rate is approximately twice the post-ChatGPT window average of $439 billion.

We test the linear-ramp assumption directly by augmenting Spec E with a quadratic time interaction, $\beta_{P_2} \cdot (t_P^2 \cdot D_{\text{post}} \cdot C^*)$. The estimated $\beta_{P_2} = -0.028$ per quarter^2 per unit C^* (sector-clustered SE 0.069, $t = -0.40$, $p = 0.69$) is small, negative, and not statistically distinguishable from zero. The implied end-of-period to window-average ratio under the quadratic specification is 1.89 against the linear-ramp value of 2.00, a five percent difference within the uncertainty of the underlying coefficients; we cannot reject the linear-ramp assumption at conventional significance levels, and the quadratic-implied end-of-2025 dollar

attribution is within rounding of the linear-implied value of $878 billion. The linear-ramp $\times 2$ conversion is therefore not an imposed functional form but an empirically supported parameterization of how the post-ChatGPT AI-attributable productivity slope evolves between 2022Q4 and 2025Q4.

We then triangulate against three additional methodologies: a BTOS-weighted attribution that uses direct firm-reported adoption to weight excess productivity, and a panel-regression attribution based on the cross-sectional γ from Verschuere and Cameron (2026a); all are reported in Table 8 on a common end-of-2025 annualized basis. We feature the DiD-based methodology as primary because it is the cross-country-comparable specification used throughout Section 6 and Appendix G, anchoring the United States to the same framework under which the cross-country aggregates are computed.

The DiD-based dollar attribution. The trend-growth decomposition translates the panel slope coefficient $\Delta\beta$ into a dollar figure in two steps. The window-average attribution is:

$$\Delta \text{ AI dollars}_{\text{window-avg}} = \Delta\beta_{\text{annualized}} \times \bar{C}^{*}_{\text{VA-weighted}} \times \text{VA}_{\text{5-sec}}$$

where $\Delta\beta_{\text{annualized}}$ is the annualized panel slope coefficient on the 5-sector cross-country-comparable specification (Section 6.1 Table 9: +7.91 for the United States, that is $\Delta\beta = +1.978$ per quarter $\times 4$),[9] $\bar{C}^{*}_{\text{VA-weighted}}$ is the value-added-weighted mean C* across the 5-sector primary panel (+0.348 for the United States), and $\text{VA}_{\text{5-sec}}$ is the 2024 nominal value added of the 5-sector panel ($15,952 billion, BEA Value Added by Industry, 2024). The product is +2.75 percentage points per year, the post-ChatGPT-window-average AI-attributable acceleration of U.S. productivity growth above the no-AI counterfactual; multiplied by the 5-sector VA, this gives $439 billion per year as the window average. The end-of-2025 run-rate doubles this:

$$\Delta \text{ AI dollars}_{\text{run-rate}} = 2 \times \Delta \text{ AI dollars}_{\text{window-avg}}$$

The factor of 2 is the ratio of the end-of-period rate to the window average under an effect

[9]The five-sector coefficient carries its own exact small-sample inference: at $G = 5$ both the wild cluster bootstrap (32 Rademacher sign patterns) and the C* permutation test (5! = 120 label assignments) are fully enumerated rather than simulated. The observed $t = 6.22$ attains the one-sided WCB attainable floor ($p = 0.031$) and ranks second of 120 permutations ($p = 0.017$); the two-sided WCB equals its attainable floor of $2/32 = 0.0625$, the most extreme value registrable at five clusters. The analytic sector-clustered SE behind $t = 6.22$ is 0.318 on the plain five-sector fit; the rolling-date implementation of Table 15 reports SE 0.319 and the window-restricted variant of Table 14 reports SE 0.327 for the same +1.978 point estimate, their auxiliary terms shifting the SE only in the second decimal (Table 18).

that grows approximately linearly across the post-ChatGPT window (near zero at 2022Q4, full strength at 2025Q4); it gives the headline $878 billion per year. The 95% CI is propagated linearly from the sector-clustered standard error on $\Delta\beta$ through the same two-step attribution. Consistent with the inference framework of Section 4.4, we note that a wild-cluster-bootstrap test-inversion interval would be wider than this analytic interval; the analytic sector-clustered CI is the primary interval-estimation tool throughout, for the reasons given in Section 4.4. The same two-step construction (window-average, then 2× for the run-rate) is applied uniformly to every economy in Section 6 and Appendix G.

The decomposition has a specific implication for how this figure should be read: the panel yields $\alpha_{\text{post}} = -0.37$ pp per quarter (common decay of pandemic-era productivity elevation, absorbed by the time fixed effect) and a positive slope on C* in the post-ChatGPT window. Without AI, the United States would not merely have smaller productivity gains post-2022Q4; it would have productivity declines across most sectors as the pandemic-era elevation decayed. The +2.75 pp/yr window average is the full amount of productivity gain above this counterfactual, and the $878 billion is the corresponding dollar flow at the end-of-2025 run-rate.[10]

Triangulation: BTOS-weighted attribution. As an independent check using direct firm-reported adoption data, we apply BTOS adoption rates to observed sectoral excess productivity, in place of the regression-estimated C*-scaled coefficient. We weight each sector's observed post-ChatGPT excess productivity above pre-2019 trend by the firm-weighted BTOS 2025-26 adoption rate for that sector. In a sector where 38% of firms report AI use, 38% of the sector's observed excess productivity above pre-2019 trend is attributed to AI. For each sector s:

$$\text{AI-attributable dollars}_s = \frac{\bar{E}_s \times \text{BTOS}_s^{\text{2025-26}}}{100} \times \text{VA}_s$$

where $\bar{E}_s$ is sector mean excess productivity in percentage points above pre-2019 trend, $\text{BTOS}_s^{\text{2025-26}}$ is the firm-weighted AI adoption rate from the 2025-26 BTOS vintage as a fraction, and VA_s is sector annual value added in 2024 USD billions from BEA. We average $\bar{E}_s$ over 2025 only (2025Q1-2025Q4) for the end-of-period run-rate, consistent with the headline framing (we label this end-of-period attribution Spec 2; Spec 1, the same attribution averaged

[10]The sectoral with-AI/without-AI decomposition shown in Figure 3 is computed on the five-sector primary panel and is therefore directly comparable to the $878 billion headline attribution. The five-sector restricted panel yields $\Delta\beta = +1.978$ per quarter; the fuller nine-sector U.S. panel yields $\Delta\beta = +2.088$ per quarter. The dollar attribution uses the five-sector coefficient throughout, since the five-sector specification is the cross-country-comparable base used for every economy.

over the full post-ChatGPT window, appears in Section 7.2). Aggregate U.S. AI-attributable GDP uplift is the sum across the nine A10 sectors. Uncertainty quantification uses the stationary block bootstrap of Politis and Romano (1994) with block lengths selected via Politis and White (2004); B = 10,000 draws per sector.

Table 7: BTOS-weighted AI-attributable U.S. GDP uplift (2025 full-year annualized)

A10 sector	$\bar{E}_s$ 2025 (pp)	BTOS 2025-26	VA 2024 ($B)	AI-attributable ($B)	95% CI ($B)
M_N Professional/Management/Admin	17.25	0.255	3,811	**168**	[155, 180]
K Finance and Insurance	22.96	0.304	2,229	**156**	[135, 176]
J Information	17.33	0.376	1,593	**104**	[91, 117]
G-I Trade/Transport/Accommodation	16.89	0.106	5,503	**99**	[87, 108]
B-E Industry	14.35	0.123	3,739	**66**	[59, 73]
O-Q Education/Health	11.18	0.225	2,544	**64**	[61, 67]
R-U Arts and Other Services	11.20	0.122	955	**13**	[12, 14]
A Agriculture	85.57	0.051	270	**12**	[11, 13]
F Construction	4.68	0.092	1,305	**6**	[5, 6]
U.S. Total			**21,950**	**$686**	**[$654, $717]**

The nine-sector panel covers $21,950B of 2024 U.S. value-added GDP, or 74.9% of the FRED GDP aggregate ($29,298B). Real Estate ($4,052B, 13.8%) is excluded due to imputed-rent measurement limitations; Government ($3,296B, 11.3%) is excluded because its VA is measured as input costs rather than output. The BTOS-weighted nine-sector approach covers 75% of GDP versus the 5-sector DiD approach's narrower 54% coverage; the larger coverage produces a directly comparable but methodologically distinct triangulation estimate.

The BTOS-weighted end-of-2025 estimate is $686 billion per year (95% CI [$654, $717] under IID bootstrap). Applying the Spec 1 block-length structure to capture longer-run autocorrelation widens the CI to approximately [$614, $757] billion.[11] The 22 percent gap between the end-of-2025 run-rate ($878B) and the BTOS-weighted end-of-2025 estimate ($686B) reflects the differing weighting of high-acceleration sectors across the two methodologies: the DiD specification places more weight on sectors where the C*-scaled slope is steepest (particularly Finance, where excess productivity rises from +15.3 pp on Spec 1 to +23.0 pp on Spec 2), while the BTOS-weighted approach uses direct adoption rates that compress sectoral differences.

The three estimates span a $642 billion to $878 billion band at the end-of-2025 reference point, with the DiD-based panel specification at the upper end and the BTOS-weighted and γ-based methods clustered in the $640–$690 billion range. We present these as separate measurements of related but distinct quantities rather than as a confirming triangulation. The DiD-based attribution is a regression-implied slope multiplied by the linear-ramp factor of 2; the BTOS-weighted figure is a level attribution using direct adoption rates; the γ-based

[11] Spec 2's short four-quarter window per sector causes the Politis-White procedure to default to block length 1, an effectively IID bootstrap. Applying the Spec 1 block-length structure (block lengths 1 to 4 across sectors) widens the CI by approximately $40B on each side. The point estimate is invariant; the wider CI is appropriate for readers concerned with longer-run autocorrelation uncertainty.

Table 8: Triangulation across estimation methodologies

Methodology	Point estimate	95% CI	Coverage
DiD trend-growth, run-rate (primary)	**$878B**	**[$602, $1,155]**	54% GDP (5-sector)
BTOS-weighted, end-of-2025 (Spec 2)	$686B	[$654, $717][a]	75% GDP (9-sector)
Cross-sectional γ (Verschuere and Cameron 2026a)	$642B	[$368, $916][b]	65% GDP

[a]IID-bootstrap interval; the autocorrelation-adjusted interval is [$614, $757], see text. [b]Propagated linearly from $\gamma_{US} = 9.37$ (SE 2.05, 95% CI [5.37, 13.37]) reported in Verschuere and Cameron (2026a) Section 4.1. The three estimates are evaluated at the same calendar reference point (end-of-2025) but measure related rather than identical quantities: the DiD figure is the regression-implied AI-attributable productivity slope multiplied by the linear-ramp factor of 2, the BTOS-weighted figure is a level attribution using observed sectoral excess productivity weighted by direct adoption, and the γ-based figure is a cross-sectional gradient on cumulative excess productivity. The DiD-based primary estimate carries the widest CI because it propagates regression-coefficient uncertainty from the panel slope through the attribution formula; the BTOS-weighted approach produces a tighter CI because direct adoption measurement removes one source of regression uncertainty. CIs from the regression-based methodologies are propagated linearly from the regression-coefficient CI rather than from the joint distribution of all attribution-formula inputs.

figure is a cross-sectional gradient on cumulative excess productivity. The three methods draw on overlapping U.S. sectoral productivity data, so their agreement does not constitute independent confirmation; what it does indicate is that no single methodological choice mechanically inflates the dollar magnitude relative to the others. The cross-method agreement should be read with appropriate caution given the shared data, and the contrast with the EU triangulation (Section 6.4), where three methods diverge across sign, is informative about the limits of cross-method agreement as evidence of identification. We use the DiD-based methodology as the headline for the cross-country-comparability reasons stated at the start of this subsection.[12]

Scope and limitations of the identification. Before extending the methodology cross-country, we flag three limitations of the U.S. identification that bound how the headline figures should be read. First, the parallel-trends diagnostic in Appendix B is conducted over four pre-period year-end quarters under a sector-clustered F-test; we supplement this with a non-overlapping pre-vs-post regime placebo (Section 7.3 and Table 14) showing that the C*-by-time slope is significantly negative in the post-COVID, pre-ChatGPT window ($\beta = -3.86$

[12]Verschuere and Cameron (2026a) report a cross-sectional γ-based estimate of $642 billion, computed with $\gamma = 9.37$ on a nine-sector functional panel covering approximately 65 percent of U.S. GDP at the 2025Q4 cutoff. That estimate and the DiD-based figures here are computed on different specifications (a cross-sectional γ estimator on nine functional sectors versus a panel difference-in-differences estimator on the five-sector cross-country-comparable panel), so they are related but not identical measures of the same underlying quantity; Verschuere and Cameron (2026a) note the same point, describing their γ-based figure as the framework's initial empirical anchor rather than a definitive measurement.

pp/qtr per unit C*, $p = 0.03$; Variant B of Table 14, Variant A is similar) and significantly positive in the post-ChatGPT window ($\beta = +1.98$ pp/qtr, $p < 0.001$), with the slope change identified at the 2022Q4 boundary ($\Delta = +5.84$ pp/qtr per unit C*, $p = 0.002$). A fuller per-quarter event study with Rambachan and Roth (2023)–style sensitivity analysis remains the appropriate next step and is reserved for future work. Second, the C* exposure measure is correlated with cognitive task content and with pre-2022 sectoral productivity (the sector fixed effects in Table 19 rise with C^*, and individual pre-period quarters occasionally show significant positive cross-sectional γ; Section 5.2), so the identifying claim is that the C*-by-time slope changed at 2022Q4, not that C* is exogenous to sectoral productivity dynamics more broadly. Third, C* is itself a constructed measure (Verschuere and Cameron 2026a Section 3) and the dollar attribution would shift under an alternative exposure construct; the companion paper reports correlations with Felten-Raj-Seamans AIOE and Eloundou et al. (2023) E2 scores that support the C* ordering, but the full dollar-attribution sensitivity to alternative exposure measures is not redone here. The pre-vs-post regime placebo strengthens the identification by showing the C*-graded slope is concentrated at the 2022Q4 boundary, but the remaining two limitations (C* correlated with pre-period cognitive task content and C* itself constructed) mean the headline magnitudes should still be read with the appropriate uncertainty that a quasi-experimental design warrants.

6. Cross-Country Results

The US shows a measurable AI productivity acceleration post-ChatGPT. Does the same mechanism operate in other advanced economies? We test on eighteen economies organized into three panels: twelve EU economies with harmonized Eurostat sectoral data, five non-EU OECD economies (United Kingdom, Japan, Canada, Australia, South Korea) with quarterly sectoral national accounts, and China with annual sectoral data from UN UNData and the National Bureau of Statistics. The structure of this section: we describe the principled sector restriction that makes productivity tests apples-to-apples with adoption tests (Section 6.1), report panel results separately for the EU pooled panel (6.2), China (6.3), and the non-EU OECD economies (6.4), summarize the dollar attribution across all economies (6.5), and document the failure of published digital-readiness composites to predict observed AI productivity outcomes (6.6).

The C* mechanism is validated cross-country in the companion methodology paper (Verschuere and Cameron 2026b). In every one of twelve EU economies tested with Eurostat 2024 sectoral AI adoption data, within-country regression of adoption on C* returns positive significant slopes with R^2 between 0.58 (Portugal) and 0.76 (Finland); the pooled cross-country slope is $\beta = +0.614$ ($t = +10.5$, $p < 0.0001$, $R^2 = 0.74$, $n = 60$). The sector-level

rank-order of AI adoption is the same in every economy, with Information the most-adopted and Construction the least, ratios of 5-19×. C* is therefore validated as a cross-country proxy for AI deployment intensity, and we use it as such throughout this section without further validation in main text.

Four findings emerge from the cross-country panel results that follow. First, EU productivity translation is positive and statistically significant on the principled five-sector restriction ($\Delta\beta = +0.614$ per quarter, $p = 0.019$, equivalent to $+2.46$ per year), at approximately 31 percent of the US annualized rate of $+7.91$ per unit of C*.[13] Second, China provides statistically significant supporting evidence on its five-sector annual panel under the China-specific 2023 treatment-date specification reflecting domestic LLM enterprise availability ($\Delta\beta = +3.80$ per year per unit of C^*, sector-clustered $p = 0.019$ referenced to $t(G-1) = t(4)$; WCB one-tailed $p = 0.032$, $G = 5$ clusters), at approximately 48 percent of the US annualized rate. Under the OECD-common 2022Q4 boundary, the China coefficient is not distinguishable from zero ($\Delta\beta = +0.67$, $p = 0.86$); the boundary-choice sensitivity is consistent with the deployment-timing argument, since Chinese enterprise generative-AI availability did not inflect until 2023 (Section 6.3). Third, non-EU OECD evidence on the quarterly panels is mixed: Australia shows a positive point estimate on the 5-sector specification ($+3.37$, $p = 0.101$) that reaches significance on the nine-sector panel once its outlier Agriculture series is excluded ($+4.22$, $p = 0.037$; Section 6.4), Japan a positive but insignificant point estimate, while the United Kingdom, Canada, and South Korea return small or negative point estimates that are not statistically distinguishable from zero, partly because several of these economies show pre-existing high-versus-low-C* divergence that the panel specification correctly does not attribute to AI. Fourth, published digital-readiness composites do not predict the cross-country pattern of AI productivity translation we observe: WIPO Global Innovation Index 2025 and OECD DESI-equivalent composites are uncorrelated with country-specific γ ($p = 0.90$ and 0.76 respectively), suggesting that infrastructure-based readiness frameworks miss what drives differential outcomes.

6.1 Cross-Country Panel DiD: The Sector-Restriction Argument

The panel specification from Section 4.3 applied country-by-country to the full 9-sector panel (Table 9) returns weak or mixed results in most EU economies, with mean $\Delta\beta$ of $+1.37$, only Germany and Sweden individually significantly positive, and Spain significantly negative. Diagnostic analysis reveals that these results are dominated by sectors where C* is a poor predictor of post-ChatGPT productivity for reasons unrelated to AI adoption.

[13] The numerical coincidence with the pooled adoption-validation slope of $+0.614$ cited above is exactly that, a coincidence: the two quantities are unrelated, one a cross-country adoption-on-C* slope in adoption-share units, the other a panel productivity DiD coefficient in percentage points per quarter.

Four sectors drive the contamination. Finance (K, with C* = 0.428) faces EU-specific banking pressures post-2020 (negative interest rate margins, restructuring, legacy NPLs); separately, Eurostat does not publish sector-level AI adoption rates for K in `isoc_eb_ain2`, so this sector cannot be included in the cross-country apples-to-apples test even if its productivity dynamics were clean. Public services (O-Q) have productivity imputed from input costs in most national accounts, mechanically suppressing measurable productivity gains. Agriculture (A) is small, weather-driven, and dominated by commodity-cycle volatility. Arts/Other (R-U) is heterogeneous and partially covered. None of these four sectors is jointly observable in Eurostat adoption data; the C* concept itself is most reliable for the five sectors where adoption is also measurable.

The principled restriction is to test the C* $\rightarrow$ productivity link only on the same five sectors where C* $\rightarrow$ adoption is validated in Verschuere and Cameron (2026b): J, M_N, G-I, B-E, F. This makes the productivity test apples-to-apples with the adoption test.

The asymmetry of the restriction is itself informative. When the same restriction is applied to the US, annualized $\Delta\beta$ moves from +8.35 (the nine-sector full panel of Table 9; +8.48 when Agriculture is also excluded, the Section 7.2 robustness convention) to +7.91 (5-sector), essentially unchanged, with the full panel significant at $p < 0.001$ and the five-sector specification at $p = 0.003$. Across the panel as a whole the picture changes substantially: under the full panel most EU economies are individually insignificant, Spain is significantly negative, and the EU mean country-level annualized $\Delta\beta$ is +1.37. Under the 5-sector restriction the EU simple-mean country-level annualized $\Delta\beta$ moves from +1.37 to +2.46, and negative point estimates attenuate toward zero in every case. Because the 5-sector EU column reports right-tailed WCB p-values, which by construction cannot certify a negative coefficient as significant, the asymmetry argument rests on this attenuation of every negative point estimate toward zero, not on an absence of significant negatives. If the restriction were arbitrary we would expect roughly symmetric effects across countries; the asymmetry, US barely moving while EU economies systematically move positive, confirms the restriction removes EU-specific contamination rather than generic noise.

6.2 EU Result

We pool the twelve EU panel economies into a single regression with country fixed effects, sector fixed effects, and time fixed effects, allowing a single C*-scaled productivity slope:

$$\text{excess}_{c,s,t} = \alpha_c + \alpha_s + \alpha_t + \Delta\alpha \cdot C_s^* \cdot D_{post} + \Delta\beta \cdot C_s^* \cdot t \cdot D_{post} + \varepsilon_{c,s,t}$$

with n = 2,640 (12 countries $\times$ 5 sectors $\times$ 44 quarters). The pooled $\Delta\beta$ is +0.614 per

quarter (= +2.46 per year), two-way clustered, p = 0.019.[14] Wild cluster bootstrap inference using 60 country-sector clusters returns p = 0.115 (two-tailed Rademacher) and p = 0.056 (right-tailed), leaving the result marginal at the 5 percent level under one-sided WCB and outside it under two-sided WCB. The pooled EU coefficient is therefore clearly significant under analytic two-way clustering but borderline under wild cluster bootstrap; we read this as cross-country evidence supportive of the headline US result rather than as independently dispositive. A pooled EU + US variant estimating separate group slopes within a common time structure returns $\Delta\beta_{US} = +1.067$ and $\Delta\beta_{EU} = +0.690$ per quarter (two-way clustered $p < 0.001$ and $p = 0.008$; Table 18); the pooled US coefficient is mechanically smaller than the stand-alone five-sector +1.978 because the shared cross-country time fixed effects absorb common post-2022 dynamics that the single-country specification attributes to the C*-scaled slope. Nine of the twelve EU countries show positive country-specific $\Delta\beta$ point estimates, of which three are individually significant at the 10 percent level under one-sided wild cluster bootstrap (Sweden, Germany, Finland); per-country inference is power-limited, as discussed below. The pooled EU annualized rate of +2.46 per unit of C* is approximately 31 percent of the US annualized rate of +7.91.

The country-by-country results behind the pooled estimate, alongside the full-panel comparison referenced in Section 6.1, are reported in Table 9. As we can see, under the 5-sector restriction: four of 19 economies clear 5 percent significance (US ***, Sweden **, Germany **, China **), with Finland marginal at the 10 percent level (p = 0.093) and Australia just outside it (p = 0.101); negative full-panel coefficients attenuate toward zero in the 5-sector specification in every case (United Kingdom from −1.16 to +0.01; Belgium from −5.06 to −0.49; Canada from −4.58 to −1.17; South Korea from −9.72 to −2.73); and Spain's significant negative coefficient (−7.72 **) reduces in magnitude to −4.00 (the 5-sector column's right-tailed p of 0.937 is uninformative for negative estimates; the 9-sector Spain coefficient is significantly negative under left-tailed WCB, Appendix B).

Table 9: Full-panel vs 5-sector $\Delta\beta$ across the cross-country panel (annualized)

Country	$\Delta\beta$_full	p	$\Delta\beta$_5sec	p	Δ	Sectors	Frequency
United States	+8.35 ***	<0.001	+7.91 ***	0.003	−0.44	9/5	Quarterly
Sweden	+10.87 **	0.011	+9.35 **	0.032	−1.52	9/5	Quarterly
Denmark	+7.29	0.157	+4.46	0.316	−2.83	9/5	Quarterly
Netherlands	+2.74	0.118	+3.74	0.121	+1.00	9/5	Quarterly
Italy	+2.08	0.571	+5.17	0.180	+3.09	9/5	Quarterly

[14] Under the canonical specification the pooled coefficient is $\Delta\beta_{EU} = +0.614$ per quarter, two-way clustered SE 0.262, $t = 2.35$, $p = 0.019$; the full inference battery is in Table 18.

Table 9 – *Continued from previous page*

Country	$\Delta\beta$_full	p	$\Delta\beta$_5sec	p	Δ	Sectors	Frequency
Finland	+2.46	0.659	+6.49 *	0.093	+4.03	9/5	Quarterly
China	+3.89	0.17	+3.80 **	0.019	−0.09	9/5	Annual
Australia	+0.17 †	0.965	+3.37	0.101	+3.20	9/5	Quarterly
Germany	+2.30 **	0.035	+2.90 **	0.031	+0.60	9/5	Quarterly
France	+3.21	0.279	+0.74	0.371	−2.47	9/5	Quarterly
Japan	+1.66	0.808	+1.48	0.817	−0.18	9/5	Quarterly
Portugal	−0.25	0.930	+0.03	0.468	+0.28	9/5	Quarterly
Austria	+0.48	0.732	+1.71	0.222	+1.23	9/5	Quarterly
United Kingdom	−1.16	0.581	+0.01	0.995	+1.17	9/5	Quarterly
Norway	−1.93	0.165	−0.64	0.713	+1.29	9/5	Quarterly
South Korea	−9.72	0.146	−2.73	0.169	+6.99	9/5	Quarterly
Canada	−4.58	0.212	−1.17	0.605	+3.41	9/5	Quarterly
Belgium	−5.06	0.332	−0.49	0.597	+4.57	9/5	Quarterly
Spain	−7.72 **	0.037	−4.00	0.937	+3.72	9/5	Quarterly
EU 12 simple mean	**+1.37**		**+2.46**		**+1.09**		

Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Inference: sector-clustered standard errors for the full-panel column; in the 5-sector column, p-values for the twelve EU economies are one-sided wild cluster bootstrap values (Rademacher sector weights), while the United States, China, and the five non-EU OECD economies use analytic sector-clustered p-values. The full panel comprises all available A10 sectors excluding Real estate (L) — nine sectors per economy — per the primary-panel convention of Section 3. † Australia's full-panel estimate is distorted by an extreme Agriculture series; excluding Agriculture (the Section 7.2 robustness convention) yields +4.22 ($p = 0.037$). $\Delta\beta$ values are annualized: quarterly slope coefficients × 4 for OECD economies (where the underlying regression uses quarterly time index); annual slope coefficient unchanged for China. The annualization places all economies on a common pp-per-year-per-unit-C* scale for cross-country comparison. EU 12 simple mean is the unweighted average of per-country $\Delta\beta$ values; the pooled-regression EU $\Delta\beta$ on quarterly data (+0.614 per quarter, equivalent to +2.46 per year) differs slightly from this simple mean because the pooled spec weights by within-country variation rather than treating countries equally.

Per-country 5-sector inference is power-limited: with five sector clusters the wild cluster bootstrap floor is $2/32 = 0.0625$ (32 unique Rademacher sign patterns) for a two-sided test and $1/32 \approx 0.031$ one-sided; the one-tailed China p of 0.032 in Section 6.3 sits at this floor within Monte Carlo error of the $B = 9{,}999$ simulation, and per-country panels have 220 observations (44 quarters × 5 sectors) rather than 396 (9 sectors × 44 quarters) for OECD or 55 (5 sectors × 11 years) for China. The pooled inference reported above is the headline cross-country EU test for this reason.

6.3 China

We extend the analysis to China using an annual sectoral productivity panel constructed from UN UNData Table 2.5 (2015-2020 constant-prices real value-added), NBS Statistical Yearbook Table 3-6 (decomposition of M_N, O-Q, R-T tertiary subgroups), and NBS Q4 advance press releases (2021-2025 sectoral real growth rates). Sectoral employment is constructed from NBS Yearbook Table 4-6 sectoral shares applied to NBS three-strata totals. Productivity is computed as value added per person employed and indexed to 2019 = 100. Validation against NBS official aggregate GDP growth produces residuals of 0.04-1.13 percentage points annually 2016-2025 (mean absolute deviation 0.45 pp).

ChatGPT was unavailable in China throughout the sample period; OpenAI services were blocked at the Great Firewall. Domestic large language models entered enterprise production with a documented mid-2023 to early-2024 timeline: Baidu Ernie 4.0 (October 2023), Alibaba Tongyi Qianwen scaling (mid-to-late 2023), ByteDance Doubao reaching enterprise scale (early 2024). The Chinese enterprise generative-AI inflection is therefore best dated to mid-to-late 2023, distinct from the 2022Q4 OECD common boundary used elsewhere in this paper. We adopt 2023 as the China-specific treatment date as the primary specification; results under the 2022Q4 OECD boundary are reported as robustness.

Under the harmonized 5-sector specification (J, M_N, G-I, B-E, F) with 2015-2018 pre-trend window and the China-specific 2023 treatment date, the pooled DiD slope coefficient is $\Delta\beta = +3.80$ (sector-clustered SE = 1.00, p = 0.019 referenced to $t(4)$) on n = 55 observations (5 sectors $\times$ 11 years). Wild cluster bootstrap inference with $G = 5$ sector clusters returns $p = 0.062$ (two-tailed Rademacher) and $p = 0.032$ (one-tailed, $B = 9{,}999$); the China result is clearly significant under analytic clustering, borderline under two-tailed WCB, and significant at the 5 percent level under one-tailed WCB. Consistent with the inference framework of Section 4.4, we interpret this as cross-country evidence supportive of the headline US finding rather than as independently dispositive; the small cluster count ($G = 5$) is at the practical floor for WCB validity. Dollar confidence intervals for China elsewhere in the paper propagate $\pm 1.96 \cdot \text{SE}$ for uniformity with the other economies; a $t(4)$ critical value (2.776) appropriate to $G = 5$ would widen [\$116B, \$365B] to approximately [\$65B, \$415B].

Under the alternative 2022Q4 OECD-common boundary, the China coefficient collapses to $\Delta\beta = +0.67$ (sector-clustered SE = 3.72, p = 0.86, WCB two-tailed p = 0.81). This is not a contradictory finding: it is exactly what the AI causal interpretation predicts. The 2022Q4 boundary *mis-dates* the Chinese treatment, because ChatGPT was unavailable in China and domestic enterprise generative-AI products did not reach scale until mid-to-late 2023. Under the wrong treatment date, treating 2022 as a post-treatment year averages

China's pre-AI productivity dynamics into the post-period mean and dilutes the C*-graded signal that materializes in 2023-2025. The sensitivity of the China coefficient to boundary choice is therefore consistent with the deployment-timing argument: where the treatment is dated correctly (2023, when domestic LLMs reached enterprise scale), the C*-graded signal is present and statistically significant; where the treatment is dated to the wrong calendar year for China specifically (2022Q4), the signal disappears. We feature the 2023 specification as primary on these economic grounds while transparently reporting the 2022Q4-boundary null result. The 9-sector specification excluding L (Real Estate, dropped because the 2021-2024 property crisis dominates real-estate productivity dynamics independently of any AI mechanism) returns $\Delta\beta = +3.89$ ($p = 0.17$); the 9-sector specification has weaker inference but the same qualitative result. We feature the 5-sector specification as the cleaner test.

The $\Delta\beta$ result is driven primarily by the J (Information & Communication) sector, where the post-2023 trajectory accelerates above the already-fast 2015-2018 pre-trend extrapolation. J productivity per worker grew approximately 94 percent over 2019-2025 (post-2022 annualized rate ~12.0 percent against a pre-2022 rate of ~11.4 percent, with acceleration concentrated in 2024-2025 as domestic LLM enterprise integration matured). M_N also accelerates modestly. The cross-sectional gradient steepens: high-C* sectors show post-treatment widening above their pre-trend extrapolations, while lower-C* sectors do not. The result captures the joint effect of multiple post-2022 mechanisms, AI deployment alongside post-COVID services recovery, post-crackdown normalization, and continued platform-economy investment, and we do not claim isolation of a pure AI mechanism.

Under the primary 2023 China-specific specification reflecting documented domestic LLM enterprise availability, China returns +3.80 percentage points per year per unit of C*, against the US annualized rate of +7.91 pp/yr (US $\Delta\beta$ of +1.978 per quarter $\times$ 4). China's primary rate is therefore approximately 48 percent of the US annualized rate. China and the US are nevertheless the two largest single-country effects in the panel, in the same order of magnitude, with the US the larger.

In the US, post-ChatGPT high-C* sectoral productivity gains coincide with employment reductions in the same sectors (output up, labor down). The Chinese pattern differs: J productivity per worker grew approximately 94 percent over 2019-2025, but Chinese tech-sector employment grew approximately 9 percent over the same period. The Chinese signature is productivity augmentation alongside continued employment growth, rather than productivity gains through labor substitution. Both produce positive value-added-per-worker results consistent with AI productivity effects, through different mechanisms.

The China case provides statistically significant supporting evidence for the C*-conditional

productivity hypothesis under harmonized methodology. Combined with the US, EU pooled, and non-EU OECD results, the pattern across the cross-country panel is consistent with AI productivity translation tracking deployment intensity: the two economies most aggressively deploying AI in the 2023-2025 window (US, China) show the largest C*-graded productivity widening; economies with smaller or slower deployment show smaller or statistically uninformative effects.

6.4 Non-EU OECD: Per-Country Results

We extend the methodology to five non-EU OECD economies with quarterly sectoral national accounts: United Kingdom (ONS), Japan (ESRI), Canada (StatCan), Australia (ABS), and South Korea (BOK with OECD annual data). The same 5-sector primary specification (J, M_N, G-I, B-E, F) and 2022Q4 treatment date are applied. We do not pool these economies because they share little structural similarity: they differ in language ecosystem, AI deployment timeline, sectoral composition, and regulatory environment. Each is reported individually.

Table 10: Non-EU OECD per-country $\Delta\beta$ (5-sector primary, annualized)

Country	$\Delta\beta$_full	p	$\Delta\beta$_5sec	p
Australia	$+0.17$ †	0.965	$+3.37$	0.101
Japan	$+1.66$	0.808	$+1.48$	0.817
United Kingdom	-1.16	0.581	$+0.01$	0.995
Canada	-4.58	0.212	-1.17	0.605
South Korea	-9.72	0.146	-2.73	0.169

Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Sector-clustered SEs. $\Delta\beta$ annualized (quarterly coefficient $\times$ 4). † Full-panel estimate distorted by an extreme Agriculture series; excluding Agriculture yields $+4.22$ ($p = 0.037$).

Per-country point estimates under the 5-sector restriction range from $+3.37$ (Australia) down to -2.73 (South Korea). Reading these results in the context of the broader cross-country panel (Section 6.1, Table 9), four economies (the United States, Germany, Sweden, and China) return individually significant positive $\Delta\beta$ at the 5-sector restriction, with Finland marginal at the 10 percent level ($p = 0.093$) and with magnitudes that scale broadly with each economy's measured AI deployment intensity. Australia, the strongest non-EU OECD case, sits just outside the boundary ($p = 0.101$) on the 5-sector specification but is significant on the nine-sector panel once its outlier Agriculture series is excluded ($+4.22$, $p = 0.037$); among the remaining four, Japan returns a positive but insignificant point estimate, while the United Kingdom, Canada, and South Korea return small or negative point estimates that are not statistically distinguishable from zero. Pre-existing high-versus-low-C* diver-

gence in Canada, the United Kingdom, and especially Japan complicates clean attribution: the panel specification correctly absorbs these pre-trends through time fixed effects, leaving smaller post-treatment $\Delta\beta$ values than the visible widening at the country aggregate level. The full-panel results are more volatile still, with three of the five economies negative under the 9-sector spec; the 5-sector restriction substantially attenuates the negative magnitudes (e.g., UK from -1.16 to $+0.01$; Korea from -9.72 to -2.73), consistent with the contamination story documented across the panel. The non-EU OECD evidence is therefore best characterized as directionally consistent with the EU and China findings, with Australia the strongest case but no economy clearing the 10 percent boundary on the 5-sector specification at current sample lengths.

6.5 Cross-Country Trend Growth Decomposition

The cross-country analysis can be extended by leveraging our framework and translating the panel-DiD coefficient into the metric most relevant to policy: trend productivity growth in percentage points per year, with and without the AI-attributable contribution. For each economy we use the fitted Spec E regression to construct two counterfactual post-ChatGPT trend growth rates: "With AI" is the realized post-ChatGPT growth, aggregated to an economy-wide CAGR using nominal value-added weights across the 5-sector primary specification (2023–2025). The AI contribution is $\Delta\beta \cdot \overline{C^*}$, where $\overline{C^*}$ is the value-added-weighted mean of C^* across the five sectors, and $\Delta\beta$ is the same Spec E panel slope that underlies the dollar attribution. "Without AI" is the realized growth minus this contribution — equivalently, the fitted model's prediction with $\Delta\beta = 0$, holding sector fixed effects, time fixed effects, and the COVID-period level shift constant, up to average regression residuals. Figure 3 displays all twenty panels (nineteen single economies plus the EU 12 pooled aggregate), each showing three bars: Without AI (the model counterfactual), With AI (realized post-ChatGPT growth), and Δ AI (the contribution).

Three findings emerge from the decomposition.

First, the United States shows the largest AI-attributable contribution in absolute terms: realized post-ChatGPT trend growth of $+5.0$ pp/yr against a Without-AI counterfactual of $+2.2$ pp/yr, an AI contribution of $+2.75$ pp/yr (window average over 2023–2025; displayed as $+2.8$ in Figure 3), approximately 56 percent of realized 5-sector trend growth. This is the same quantity that underlies the dollar attribution in Section 5.4 and Table 8 — the panel slope multiplied by VA-weighted mean C* — so the percentage and dollar views are two expressions of a single estimate.

Second, in many advanced economies the "Without AI" counterfactual is negative or near

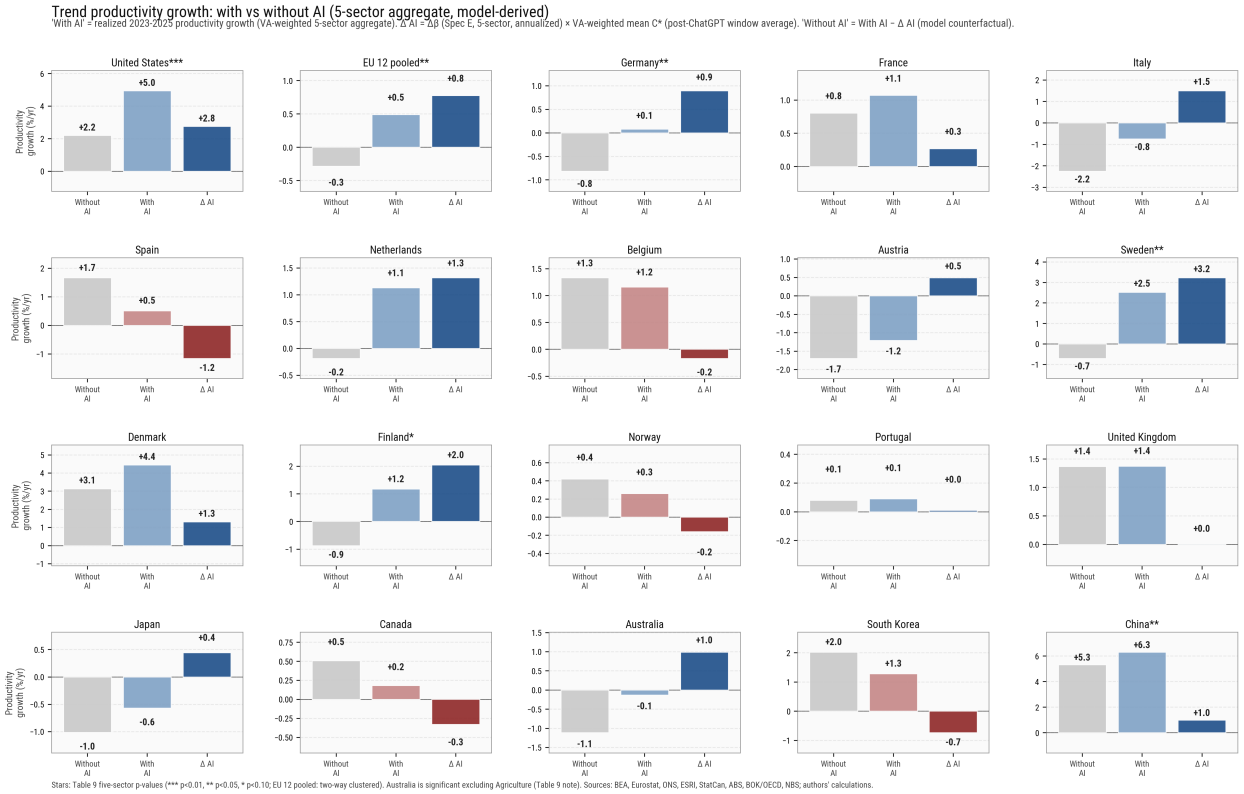


Figure 3: Trend Productivity Growth With vs Without AI, 19 Economies + EU Pooled, 2023–2025

zero, meaning the realized post-ChatGPT productivity growth is largely or entirely AI-attributable. Of the twelve EU economies, six (Austria, Finland, Germany, Italy, the Netherlands, Sweden) have negative Without-AI counterfactuals; the remaining six (Belgium, Denmark, France, Norway, Portugal, Spain) have positive counterfactuals, with Belgium, Norway, and Spain pairing them with negative or near-zero Δ AI. Japan and Australia also show negative Without-AI counterfactuals offset by positive Δ AI. The EU 12 pooled decomposition: Without AI −0.3 pp/yr, With AI +0.5 pp/yr, Δ AI +0.78 pp/yr. The cross-country evidence supports an interpretation in which AI is preventing what would otherwise be a continued productivity stagnation regime in half of the EU 12, reversing rather than augmenting the underlying growth rate.

Third, China shows a large realized productivity rate (+6.3 pp/yr) with a modest AI contribution (+1.0 pp/yr) on top of an already-high counterfactual (+5.3 pp/yr). The Chinese pattern is qualitatively different from both the US and the EU: most of China's post-ChatGPT productivity growth comes from non-AI sources (continued catch-up dynamics), with AI providing a marginal additional lift. The proportional AI contribution to Chinese trend growth is approximately 16 percent, versus approximately 56 percent for the US.

Finally, at the aggregate level our model shows AI currently generating close to $1.26 trillion per year at the end-of-2025 run-rate, raising aggregate trend productivity growth on the combined U.S.–EU 12–China 5-sector base from +2.6 percent per year (counterfactual) to +4.3 percent per year (realized). The full attribution is reported in Appendix G, where each country's Δ AI is translated into 5-sector nominal value-added terms. The United States, EU 12 pooled, and China together account for $37.1 trillion in 5-sector value added. The United States accounts for $878 billion of the total (approximately 70 percent), with China at $240 billion and the EU 12 pooled aggregate at $138 billion, all at the end-of-2025 run-rate. Across the full set of nineteen single economies in the panel (United States, twelve EU economies individually, five non-EU OECD economies, and China; $44.6 trillion in combined 5-sector VA), Δ AI sums to approximately +$1.26 trillion per year at the run-rate. Several individual economies (notably Germany, France, Spain, and Norway) show negative point estimates under the Eurostat-adoption-weighted attribution reported in the EU triangulation below, reflecting that their high-exposure sectors remain below pre-ChatGPT productivity trend in absolute terms even as the DiD specification identifies a positive relative slope. We emphasize that all dollar magnitudes in this paper are computed on the 5-sector primary panel base; aggregate dollar flows including the broader sector mix would be larger.

Triangulation for the EU 12. Just as the U.S. headline is triangulated against direct-survey and cross-sectional methods in Section 5.4, the EU 12 pooled figure can be checked against two independent estimates. The cross-sectional γ-based attribution applies the framework's U.S. methodology (Verschuere and Cameron 2026a) to the EU 12 pooled cross-section, running $\text{ExcessGrowth}_s = \gamma \times C_s^*$ origin-forced over the nine A10 sectors at 2025Q4. The estimated $\gamma_{\text{EU}} = +4.11$ is positive but very imprecise ($R^2 = 0.03$, $n = 9$), reflecting that the EU cross-section is dominated by two non-AI structural outliers, with Real Estate and Finance off-trend for reasons unrelated to AI (a real-estate productivity bounce and incomplete post-pandemic banking-sector recovery, respectively). The implied dollar attribution is approximately +$136 billion per year with a wide 95% CI that includes zero. The adoption-based attribution uses Eurostat's `isoc_eb_ain2` AI-adoption survey, the EU analogue of the U.S. Census BTOS supplement: a direct, firm-reported measure of sectoral AI deployment. Weighting each EU sector's observed post-ChatGPT excess productivity by its Eurostat-measured adoption rate, in place of the regression-estimated C^*-scaled coefficient, and scaling to the end-of-2025 annualized rate on the same basis as the DiD run-rate, yields an adoption-based EU 12 estimate of approximately −$101 billion per year.

The EU triangulation does not converge on a tight headline as the U.S. triangulation does, and the divergence is more diagnostic of what cross-method evidence on this question typ-

Table 11: Triangulation across estimation methodologies, EU 12 pooled

Methodology	Point estimate	95% CI	Coverage
DiD trend-growth, run-rate (primary)	**+$138B**	**[+$23, +$254]**	54% GDP (5-sector)
Cross-sectional γ[a]	+$136B	[−$378, +$651]	5-sector VA, 9-sector γ
Eurostat-adoption-weighted, end-of-2025[b]	−$101B	[−$110, −$93]	5-sector

[a]Cross-sectional γ-based estimator parallel to Verschuere and Cameron (2026a) for the United States: $\text{ExcessGrowth}_s = \gamma \times C_s^*$ on the EU 12 pooled 2025Q4 cross-section, origin-forced. Estimated $\gamma_{\text{EU}} = +4.11$ (SE 7.92, $R^2 = 0.03$, n = 9), with the regression dominated by two off-gradient outliers (Real Estate and Finance, discussed below). The point estimate is positive but the 95% CI is very wide and includes zero, reflecting both small sample (n = 9) and the structural non-AI shocks that dominate the EU post-pandemic cross-section. Dollar attribution = $\gamma_{\text{EU}} \times \bar{C}^*_{\text{9-sec}} \times \text{VA}_{\text{EU 5-sec}} = 4.114 \times 0.372 \times \$8{,}909\text{B}/100 \approx \$136\text{B}$; γ_{EU} is measured on the A10 cumulative-excess basis (Section 5.2 footnote), origin-forced with no intercept and the pre-trend fitted on 2015Q1–2018Q4 (the cross-country γ convention, as in the China specification), with the 95% CI on γ of $[-11.4, +19.6]$ propagated linearly to the dollar interval. [b]Per-country bootstrap CIs from the replication-package attribution routine on each of the 12 EU economies (block bootstrap over quarterly excess productivity series), aggregated to the EU 12 pooled total assuming independence across countries; the pooled standard error is the square root of the sum of country-level variances, then scaled ×2 to convert from the post-ChatGPT window average to the end-of-2025 annualized run-rate. All three estimates are end-of-2025 annualized rates. The Eurostat-adoption-weighted estimate is the EU analogue of the U.S. BTOS-weighted estimate in Table 8: it weights observed sectoral excess productivity by direct firm-reported adoption rather than by the regression-estimated C^* gradient. Unlike the U.S. case, the three EU methods do not converge on a tight range; this divergence is discussed in the text.

ically looks like than the U.S. agreement is. The DiD-based estimate (+$138B) and the cross-sectional γ-based estimate (+$136B) are positive and agree in sign, with the γ-based estimate having a very wide CI that includes zero; the Eurostat-adoption-weighted estimate is negative (−$101B). The honest reading is that the cross-sectional methodology has limited resolving power in the EU context. The cross-sectional regression yields $R^2 = 0.03$, with the C^* gradient effectively obscured by two off-gradient outliers (Real Estate and Finance) that move for reasons unrelated to AI: a European real-estate productivity bounce in moderate-C^* Real Estate, and ongoing post-pandemic balance-sheet repair in high-C^* Finance. These two outliers roughly cancel out the cross-sectional gradient, leaving the γ-based estimator without statistical power to identify the AI signal in the EU. The DiD specification, by contrast, can identify the C^* slope within the EU by absorbing sector-level shocks into sector fixed effects; this is why the panel-DiD methodology is the appropriate primary specification for the cross-country analysis. The negative Eurostat-adoption-weighted figure reflects, in *absolute* terms, that most EU sectors remain below their pre-ChatGPT productivity trend; weighting below-trend excess productivity by adoption produces a negative dollar figure. The U.S. does not show this absolute-versus-relative divergence because U.S. sectors are above trend in absolute terms; the EU is the case where the relative-versus-absolute distinction binds.

Read this way, the EU case is the one that exposes the underlying methodological friction: the three approaches measure related but distinct objects, and they happen to agree in the United States and disagree in Europe. We report the DiD-based run-rate as the primary EU figure for consistency with the U.S. headline and the rest of the cross-country analysis, while flagging that the cross-sectional methodology yields a wide, imprecise estimate and that the adoption-weighted reading places the EU's absolute AI-attributable flow near zero or slightly negative.

Figure 4 shows the sector-level cross-sections underlying the γ-based estimates for each of the 20 economies in our panel, plotted on the window-average basis. The attribution γ's of Tables 8 and 11 are 2025Q4-*endpoint* objects (Table 8's U.S. γ on the 2026a paper scale, Table 11's EU γ on the figure's A10 scale); for the EU 12 pooled cross-section the endpoint $\gamma = +4.11$ is approximately twice the window-average $+2.07$ shown here, as a post-2022 linear ramp implies. The visual evidence reinforces the table-level finding. The United States panel exhibits the canonical pattern: a tight upward slope from Construction and Industry at the low-C^* end through Information and Professional Services at the high-C^* end, with $R^2 = 0.83$. The EU 12 pooled panel, by contrast, shows two large off-gradient outliers — Real Estate above the line and Finance well below — that pull the regression nearly flat ($R^2 = 0.01$). The country-level panels reveal substantial heterogeneity: some EU economies (notably Italy, Portugal, Sweden) show clean positive gradients consistent with the U.S. pattern; others (Germany, Finland, Norway) show negative or flat slopes driven by sector-specific shocks rather than the C^* ordering. The non-EU 5 panel shows mostly weak or negative gradients, with Japan as a clean positive exception. China's panel, on annual data with limited resolution, shows a weak positive slope. The cross-country dispersion in panel-level γ estimates illustrates why the panel-DiD specification is the appropriate primary framework for cross-country comparison: by absorbing sector-level shocks into sector fixed effects, the DiD identifies the C^* slope within each economy from the time variation rather than the contaminated cross-section.

As we show the AI effect is already observable across advanced economies; the largest AI-attributable contributions to trend productivity growth are concentrated in the United States, Sweden, Finland, Italy, the Netherlands, and Denmark (per-GDP rates), with the United States dominating in absolute terms. The decomposition surfaces a meaningful policy fact: in much of advanced Europe and East Asia, AI is the difference between trend productivity stagnation and modest positive growth.

6.6 Composite Predictor Failure

Published digital-readiness composites are widely used as proxies for AI deployment readiness, on the premise that economies scoring high on infrastructure-and-skills metrics should be expected to translate AI capability into productivity gains more readily. The empirical pattern in our cross-country panel does not support this premise.

We compute the Pearson and Spearman correlations between each economy's published composite rank and its country-specific γ (the cross-sectional gradient of post-ChatGPT excess productivity on C*) across the 17 OECD economies in our panel with both a composite rank and identifiable γ. Two composites are tested: WIPO Global Innovation Index 2025 (taken as published) and an OECD DESI-equivalent composite constructed from eight indicators reported in OECD (2024) including connectivity, digital skills, integration of digital technology, digital public services, cybersecurity, data policy, and digital government, standardized to mean zero and unit variance and averaged.

Table 12: Cross-country composite predictor performance

Composite	Pearson r (composite rank, γ)	Spearman ρ	Bootstrap 95% CI on ρ
WIPO GII 2025	$+0.055$ ($p = 0.834$)	$+0.12$ ($p = 0.66$)	$[-0.51, +0.65]$
OECD DESI-equivalent	$+0.081$ ($p = 0.756$)	$+0.09$ ($p = 0.73$)	$[-0.54, +0.62]$

Neither composite achieves conventional significance, and the Spearman bootstrap confidence intervals each span values from substantially negative to substantially positive. The composites do not predict which economies translate AI capability into measured productivity outcomes. The composite-rank ordering does not align with observed γ: for example, Korea and the United Kingdom rank high on both composites but show small or negative γ, while Italy and Portugal rank lower but show positive γ. The mismatch suggests that infrastructure-based readiness frameworks are not capturing the variation that drives observed AI productivity translation.

The implication for AI-readiness frameworks is direct: composites built from infrastructure metrics without empirical validation against productivity outcomes do not inform expected productivity translation. Productivity-grounded measurement, combining task-level capability (C*), observed adoption, and sectoral productivity outcomes, provides a more reliable framework for ongoing monitoring of AI's macroeconomic contribution. The policy implications are developed in Section 8.

7. Robustness and Extensions

7.1 Secondary Alternative-Mechanism Tests

Four additional alternative mechanisms for the post-ChatGPT US signature are tested and reported here; the cost-of-capital alternative, the most common of all, is treated in Section 5.3 with full results in Appendix C.8. Each identifies a feature of the observed pattern it would need to reproduce and the test applied; none produces the full joint pattern. Supporting calculations are in Appendix C.

Table 13: Summary of secondary alternative mechanisms tested

Alternative	Predicted pattern	Test	Result
Mean reversion from pandemic elevation	Post-2022Q4 = continued trend from pandemic	Compare pandemic-peak vs end-of-sample levels; regress post-ChatGPT increment on pandemic excess and on C*	Excess level +7.3 pp at the 2022Q3 pandemic peak vs +16.1 pp by 2025Q4 (excl. Agriculture): continued divergence, not reversion. The post-ChatGPT increment follows C* (R^2 = 0.71) and is at most weakly related to pandemic-era excess (r = −0.39, indistinguishable from zero at n = 8; App. C.5)
Jobless-recovery dynamics	High-C* sectors had larger pandemic VA drops, employment cuts, slower recoveries	Regress each of three measures on C* across US sectors	All three slopes insignificant on the full nine-sector cross-section (p = 0.30, 0.57, 0.41); the VA-drop and employment-cut slopes are in fact positively signed (high-C* sectors had smaller pandemic drops) — the opposite of the preconditions the alternative requires
Post-recession dynamics (2008 placebo)	Deep recessions produce C*-scaled patterns in recovery	Apply the γ_t estimator and the $\Delta\beta$ slope test to the 2004-2014 window	Post-2008 recovery mean γ_t = −1.6 vs post-ChatGPT +9.7 (paper scale, excl. Agriculture), an 11.3 pp gap in the opposite direction, with the pre-recession baseline at +0.01. No C*-scaled slope break: $\Delta\beta$ = −0.85 (p = 0.34) vs +2.088 (p = 0.0002) post-ChatGPT
Date placebo tests	Specification identifies generic post-period features	Re-estimate with placebo treatment dates 2018Q4 and 2020Q4	Placebo $\Delta\beta$ = −1.52 (two-way p = 0.34) and −0.78 (two-way p = 0.33; OLS p = 0.035, see text) vs actual +2.088 (p < 0.001)

Each of these mechanisms predicts a pattern at odds with the observed data. The 2008 placebo is the most demanding of these tests: if post-recession dynamics absent AI were the operative mechanism, the 2008-2014 recovery window should exhibit the C*-scaled slope break we observe post-ChatGPT. It does not: the slope test returns $\Delta\beta = -0.85$ (p = 0.34) on the 2009-2014 window versus +2.088 (p = 0.0002) post-ChatGPT, and the level gradient is mildly negative across the recovery (mean event-study $\gamma_t = -1.6$, paper scale, against a pre-recession baseline of +0.01) versus +9.7 post-ChatGPT, the same event-study object on the same scale (the origin-forced window-average gradients of Figure 4 are a different object, on the A10 scale) — an 11.3 pp gap in the opposite direction. The date placebo tests similarly bear on the possibility that the specification mechanically identifies any post-2022 period. Placebo $\Delta\beta$ at 2018Q4 and 2020Q4 are small and signed against the direction of the actual effect (-1.52 and -0.78 respectively). Under two-way clustered inference, the appropriate test for placebos given our methodology choices (Section 4.4), both placebos are clearly insignificant (p = 0.34 and 0.33). We note for transparency that under unadjusted OLS, the 2020Q4 placebo would be borderline-significant at conventional levels (p = 0.035), and the 2018Q4 placebo would be borderline (p = 0.064). OLS ignores within-sector serial correlation and understates the placebo standard errors: the two-way clustered SEs are roughly twice the OLS SEs for both placebos, and the two-way clustered p-values are the appropriate placebo inference under the framework of Section 4.4. Under those methods the headline remains highly significant (WCB p = 0.016, permutation p = 0.006) while the placebos cannot be tested against a one-sided alternative because their point estimates are negative. The full per-method placebo p-values are shown in Table 18, alongside the headline results for direct comparison.

7.2 Specification Robustness

The primary panel specification (Section 4.3) is one of several defensible specifications. We report four additional specification robustness checks to verify the central result is not specification-specific.

Fully-interacted specification. Allowing sector-specific pre- and post-period level shifts as separate parameters (Spec A in Appendix A) yields $\Delta\beta = +3.381$ on the primary outcome (vs +2.088 under the primary specification), with permutation p = 0.0013 (rank 13 of 10,001). Both specifications reach the same qualitative verdict, but they differ quantitatively. The economic interpretation of the difference is the following. Spec E absorbs a common post-2022Q4 level decay of -0.37 pp/quarter (common to all sectors, capturing pandemic-elevation decay, calendar effects, and other one-time shocks) into the time fixed effect. The slope coefficient $\Delta\beta$ then captures the *diffusion rate* of C^*-scaled productivity gains: the

rate at which the productivity gap between high-C^* and low-C^* sectors widens over time, separately from any common one-time post-period level shift. Spec A by contrast allows the slope coefficient to absorb part of what Spec E attributes to the common level shift, because the level shift is parameterized at the sector level rather than as a single scalar. For an attribution exercise asking "what is AI's current diffusion-rate contribution to productivity growth," the Spec E slope object is the correct measurement. Spec A is informative as a verification that the headline survives even when sector-specific level shifts are allowed to soak up variance that could otherwise be misattributed to slope; it serves as a robustness check, not a competing primary specification. For the dollar attribution, the coefficient and the value-added base must match: the nine-sector Spec A coefficient (+3.381) cannot be applied to the five-sector attribution base of Section 5.4. Re-estimated on the five-sector panel itself, Spec A yields $\beta^{\text{post}} = +1.955$ per quarter (OLS SE 0.371), and the matched-base end-of-2025 run-rate attribution is approximately $868 billion per year, with 95% CI [$545B, $1,191B] — within roughly one percent of the headline Spec E attribution of $878 billion. The dollar attribution is therefore essentially invariant to the choice between the two specifications once coefficient and base are matched: allowing sector-specific level shifts to soak up variance that could otherwise be misattributed to slope moves the headline by $10 billion. The primary Spec E attribution remains our preferred reading because it isolates the diffusion-rate object rather than the combined level-and-slope object.

Leave-one-out sector robustness. Re-estimating the primary specification dropping each A10 sector in turn, the coefficient on the C* slope interaction ranges from +1.91 to +2.39 across the nine leave-one-out subsamples (all remain statistically significant at $p < 0.01$). No individual sector drives the result; the largest single influence is Finance (K), whose exclusion reduces $\Delta\beta$ by 0.18 to +1.91. Full leave-one-out table is in Appendix B.

Agriculture inclusion. Section 5.2 excludes Agriculture from its decomposition tables and cross-sectional correlations. The justification is that US Agriculture displays anomalous excess productivity readings during 2020-2022Q3, reaching +94 percentage points above pre-2019 trend by 2022Q3 and driven almost entirely by real value-added movements that track the international agricultural commodity price cycle rather than productivity dynamics. Employment in Agriculture was essentially flat throughout. The measured "gain" is consistent with imperfect deflation during commodity-price shocks: when output prices surge faster than input prices, sector-specific deflators may not fully separate prices from quantities. A genuine +80 percent productivity gain over two years is also inconsistent with Agriculture's long-run trajectory: USDA ERS data show average productivity growth of 1.46 percent annually (1948-2021) with no precedent for step changes of this magnitude (USDA Economic

Research Service 2024).

The empirical substance of the paper is robust to Agriculture's inclusion. The US panel $\Delta\beta$ is $+2.088$ in the full 9-sector sample and $+2.12$ excluding Agriculture; the pandemic-window $\Delta\beta$ collapses from -8.01 to -3.54 (about 56 percent reduction in absolute magnitude) when Agriculture is excluded, confirming that Agriculture contributes to but does not solely drive the pandemic-era pattern. The BTOS-weighted dollar attribution on the window-average specification (Spec 1) is \$556.2B with Agriculture included and \$545.1B without, with Agriculture contributing only \$11.1B (2.0 percent of the total); both estimates fall inside the 95% confidence interval of the included-Agriculture attribution.

Block-length sensitivity for dollar attribution. On the Spec 1 window-average specification (\$556B point estimate), the bootstrap standard deviation ranges from \$17.7B to \$22.0B across Politis-White automatic selection and fixed block lengths of 1, 2, 3, and 4 quarters, with 95% CI width ranging from \$70.0B to \$85.4B. The point estimate is invariant across all specifications. For Spec 2 (end-of-2025 \$686B BTOS-weighted estimate, T=4 quarters per sector), the Politis-White procedure defaults to block length 1 and the bootstrap is effectively IID; the Spec 1 block-length comparison indicates a wider CI (~\$40B wider) would result from a longer series with proper autocorrelation handling.

7.3 Pre-vs-Post Regime Placebo Test

A common identification challenge for any post-event DiD is to demonstrate that the estimated slope is identified by the post-treatment regime rather than by some pre-existing trend present throughout the post-COVID period. The cleanest test compares the C*-by-time slope in a non-overlapping pre-ChatGPT window against the same slope in the post-ChatGPT window. We run two variants: strict-symmetry (eight quarters pre versus eight quarters post, 2020Q4–2022Q3 versus 2022Q4–2024Q3) and full-power (the entire clean post-pandemic regime, 2021Q1–2022Q3, versus the full post-ChatGPT window 2022Q4–2025Q4). In each variant we fit $\text{excess}_{s,t} = \alpha_s + \alpha_t + \beta \cdot (t \cdot C^*) + \varepsilon$ on the chosen window, where t counts quarters since the window start.

The PRE-window slope is significantly *negative* in both variants (approximately -3.9 pp/qtr per unit C*, $p = 0.03$), and the POST-window slope is significantly *positive* (approximately $+2.0$ pp/qtr per unit C*, $p < 0.001$). The slope change across the 2022Q4 boundary is approximately $+5.9$ pp/qtr per unit C* with $p < 0.003$ in both variants.

This sign-flip is the regression-version of the regime-separation result already documented qualitatively in Section 5.2: the cross-sectional correlation between sectoral productivity change and C* moves from $r = -0.46$ during the pandemic window to $r = +0.86$ in the

Table 14: Pre-vs-post regime placebo: slope of $C^* \times t$ in non-overlapping windows.

Variant	Window	β /qtr	SE	t	p
A. Strict symmetry (8q vs 8q)	PRE (2020Q4–2022Q3)	−3.891	1.707	−2.279	0.031
	POST (2022Q4–2024Q3)	+2.058	0.513	+4.014	0.000
B. Full-power (7q vs 13q)	PRE (2021Q1–2022Q3)	−3.861	1.641	−2.352	0.028
	POST (2022Q4–2025Q4)	+1.978	0.327	+6.051	0.000
Variant A: Δ(POST − PRE)		+5.949	1.782	+3.337	0.003
Variant B: Δ(POST − PRE)		+5.839	1.674	+3.488	0.002

Slope β is the within-window coefficient on $t \cdot C^*$ from a panel regression with sector and time fixed effects, units of pp per quarter per unit C^*. SE is sector-clustered (CR1). The Δ(POST − PRE) statistic tests equality of slopes across the two non-overlapping windows under the assumption of independence; the joint standard error is the square root of the sum of within-window variances. The PRE slope is significantly *negative* in both variants, indicating that the post-COVID, pre-ChatGPT period exhibited high-C^* sectors *underperforming* low-C^* sectors. The POST slope is significantly *positive* in both variants. The slope change at 2022Q4 is +5.84 to +5.95 pp per quarter per unit C^*, with $p < 0.003$ in both variants.

post-ChatGPT window, and the per-quarter γ time series in Section 5.2 traces the same boundary at 2022Q4. The two diagnostics measure the same phenomenon: the C^*-graded productivity gradient exists in opposite signs across the two regimes, and changes sign at the ChatGPT release. This cross-window test expresses the finding as a single test statistic on the difference of slopes between two non-overlapping windows, complementing Section 5.2's correlation and event-study diagnostics with an explicit cross-window inference. The economic interpretation is given in Section 5.2: during the post-COVID, pre-ChatGPT period, high-C^* sectors were absorbing the tech-employment hangover from 2020–2022 over-hiring (Information added employment faster than output through 2022Q3, mechanically suppressing productivity), while low-C^* sectors benefited from pandemic-cleansing dynamics (excess-layoff productivity gains in Industry, Trade/Transport, and Education/Health). These two patterns generated the negative C^*-by-time slope in the pre-window. The post-ChatGPT gradient reverses these patterns: high-C^* sectors deliver large output gains combined with employment reductions, while low-C^* sectors normalize to trend. The post-ChatGPT C^*-graded productivity acceleration is therefore identified by the post-treatment regime change, not by a continuation of post-COVID dynamics; the negative PRE-window slope is itself a substantive finding that supports rather than undermines the identification.

7.4 Rolling Treatment-Date Robustness

A natural concern with any post-event DiD is that the result depends on the specific treatment date chosen, particularly when the "true" inflection point is not directly observable. ChatGPT's late-November 2022 public release is the canonical date used throughout the paper, but firm adoption became measurable in BTOS only from 2023 onwards. We test

sensitivity by re-fitting the full Spec E specification with treatment dates shifted across the seven quarters from 2022Q1 through 2023Q3.

Table 15: Rolling-treatment-date robustness: US Spec E with treatment quarter varied.

Treatment date	Post-T (q)	β_P /qtr	SE	t	p	Run-rate $B
2022Q1	16	+1.650	0.311	+5.30	<0.001	+733
2022Q2	15	+1.873	0.341	+5.50	<0.001	+832
2022Q3	14	+1.947	0.325	+6.00	<0.001	+865
2022Q4 (canonical)	13	**+1.978**	**0.319**	**+6.20**	**<0.001**	**+878**
2023Q1	12	+1.910	0.310	+6.16	<0.001	+848
2023Q2	11	+1.911	0.339	+5.64	<0.001	+849
2023Q3	10	+1.963	0.413	+4.76	<0.001	+872

β_P is the Spec E coefficient on $C_s^* \cdot t_P \cdot D_{\text{post}}$ with sector-clustered (CR1) standard errors and the COVID indicator D_{covid} shrunk to end one quarter before the treatment date for non-overlap. Post-T is the number of quarters from treatment to end-of-sample (2025Q4). Run-rate $B is the implied end-of-2025 annualized dollar attribution: $2 \times \beta_P \times 4 \times \bar{C}^* \times \text{VA}_{\text{5-sec}}/100$ with $\bar{C}^* = 0.348$ and VA = $15,952B. All seven specifications are highly significant ($p < 0.001$); the run-rate dollar attribution ranges from $733B (2022Q1, the earliest tested date) to $878B (canonical 2022Q4), and forward-shifted dates 2023Q1–2023Q3 produce run-rates within $30B ($848B–$872B) of the canonical value.

The canonical 2022Q4 specification produces the largest β_P and the highest t-statistic in the table, with neighboring dates producing slopes that rise monotonically up to 2022Q4 and sit slightly lower at all dates moving away from 2022Q4 in either direction. The pattern is informative on two questions that the panel-DiD literature raises about post-event specifications.

Is the treatment date well-identified? The data places the maximum at 2022Q4 without prior imposition. The slope rises monotonically from $\beta_P = +1.65$ at 2022Q1 to +1.98 at 2022Q4, then declines and plateaus at +1.91 to +1.96 across 2023Q1 through 2023Q3. The retention of statistical significance at earlier treatment dates reflects a mechanical feature of the panel-DiD specification rather than evidence against identification: under an earlier treatment date, the post-window includes both pre- and post-ChatGPT quarters, so the slope estimate is a length-adjusted average that incorporates the bulk of the actual post-2022Q4 acceleration regardless of where the indicator is placed. The cumulative excess productivity at end-of-sample ($\beta_P \times T_{\text{post}}$) is of the same order across treatment dates: $1.65 \times 16 = 26.4$pp at 2022Q1 vs $1.98 \times 13 = 25.7$pp at 2022Q4 vs $1.96 \times 10 = 19.6$pp at 2023Q3. The same measured acceleration is being parameterized as either a flatter slope over a longer run or a steeper slope over a shorter run; the maximum point of β_P identifies the inflection date empirically.

Why is the forward direction not more sensitive? Treatment-date sensitivity can be a serious concern in this class of specification; the China analysis in Section 6.3 illustrates this directly, with the $\Delta\beta$ point estimate changing from +0.67 to +3.80 when the treatment date is shifted from 2022Q4 to 2023. The U.S. result does not exhibit this sensitivity: the run-rate dollar attribution at treatment dates 2023Q1, 2023Q2, and 2023Q3 is \$848B, \$849B, and \$872B respectively, all within 4 percent of the canonical \$878B. The difference with China reflects the difference in identifying variation: the China specification operates on annual data with only 2-3 post-treatment observations under the canonical date, so date shifts of one to two quarters can change which observations enter the post-window substantially; the U.S. specification operates on quarterly data with 12-13 post-treatment observations under the canonical date, so date shifts of one to two quarters change the slope estimate only at the margin. The run-rate range across the full ±3-quarter sensitivity in the U.S. (\$733B to \$878B, with the forward-shifted subset \$848B to \$878B) is the appropriate stated uncertainty around the headline under treatment-date robustness.

7.5 Falsifiers for Future Data

The interpretation developed in this paper makes predictions about the relationship between AI adoption and measured productivity that can be tested against data not yet available. We state six falsifiers: empirical patterns that, if observed in 2026-2027 data, would be inconsistent with our interpretation. They are stated here in advance of the 2026–2027 data that can test them.

Decoupling of productivity and C*. If post-2026 productivity growth in high-C* US sectors decelerates sharply while low-C* sectors continue at trend, the C*-scaled interpretation weakens. The AI mechanism predicts that the C* gradient in observed productivity persists or widens as adoption deepens, not that it closes as high-C* sectors plateau.

Decoupling of adoption and productivity. If BTOS 2027-28 adoption rates continue rising in high-C* sectors while productivity in those sectors does not continue to accelerate, the productivity-adoption link weakens. A genuine AI-productivity relationship should produce continued productivity gains as adoption deepens toward saturation.

Appearance of the signature where adoption is low. If the panel specification returns a C*-scaled productivity signature in a country where BTOS-equivalent adoption remains low, the BTOS-C* equivalence underlying the paper's interpretation is falsified. The signature should appear where adoption is high and not appear where it is low.

Reversal of the employment response. If high-C* US sectors begin adding employment at rates comparable to low-C* sectors in 2026-2027 while productivity in high-C* sectors

continues to grow, the labor-substitution signature weakens. AI adoption predicts the joint pattern of output-up and labor-down; divergence would require reassessing the operative mechanism.

EU translation does not strengthen as adoption rises. The EU buffering interpretation developed in Sections 6.5 and 8 (low-adoption sectors continuing to decline while high-adoption sectors hold flat) implies that as EU adoption deepens through 2026-2027, the high-adoption sectors should begin showing absolute productivity gains rather than just relative resilience. If EU adoption continues to rise but high-C* productivity remains flat, the buffering interpretation needs revisiting and the operative mechanism may differ from the US substitution pattern in ways the current data cannot distinguish.

China's J-sector productivity acceleration reverses. The China interpretation in Section 6.3 rests on the J (Information & Communication) sector showing post-2023 productivity acceleration above its already-fast pre-2018 trend, alongside continued employment growth (the augmentation signature). If the 2024-2025 J productivity acceleration reverses in 2026-2027 data, either through productivity deceleration or J-sector employment reduction, the augmentation interpretation weakens and the post-2022 result becomes more plausibly attributable to post-COVID services recovery and post-crackdown normalization rather than AI deployment.

These are not the only possible falsifiers, but they are specific and testable against forthcoming data.

8. Policy Implications

The advanced economies entered the post-pandemic period with three structural constraints binding simultaneously. Inflation above target since 2021 left central banks tightening into elevated debt and a slowing labor force, with little space for the conventional response of stimulus through accommodation. Productivity growth, the third leg of the standard policy framework, had been weak across most of the developed world for over a decade, the OECD productivity-growth slowdown that motivates the Acemoglu (2024) projections this paper has revisited. Against that backdrop, the magnitude documented here (approximately +2.75 percentage points per year of AI-attributable U.S. trend productivity growth, materially smaller but still positive in most of advanced Europe and East Asia) represents a development of consequence for each leg of the framework. We discuss the implications across three policy domains, and close with the implication for the AI-readiness measurement infrastructure on which much current policy guidance is built.

The first implication concerns monetary policy. A productivity acceleration of the magni-

tude estimated here is, in standard New Keynesian terms, a positive supply shock: it raises potential output without raising aggregate demand, and the appropriate response is policy accommodation rather than restriction. The asymmetry of the result across economies is therefore directly relevant to the relative monetary stances of the Federal Reserve, ECB, and PBoC. In the United States, AI is contributing approximately 3.0 percent of nominal GDP per year to potential output (the headline AI-attributable productivity flow of $878 billion divided by U.S. 2024 nominal GDP of $29.3 trillion, FRED), a substantially larger supply shock than the typical secular-stagnation supply shocks that have featured in Fed deliberations over the past decade. Output-gap measures and Phillips-curve specifications that do not incorporate this expansion will systematically understate slack and bias the policy stance toward overtightening. The bias operates in the opposite direction in the EU 12 panel, where the AI contribution at approximately 0.8 percent of GDP (the $138 billion flow against EU 12 aggregate nominal GDP of roughly $17 trillion) is meaningful but materially smaller; the implication is that the productivity-driven case for policy accommodation is stronger in the U.S. than in the euro area, even though both face common cyclical pressures. In China, the additional ~1.2 percent of GDP contribution (the $240 billion flow against China's 2025 nominal GDP of approximately $19.6 trillion) arrives on top of an already-strong non-AI growth path, with correspondingly different implications for the policy mix.[15] The cross-country asymmetry documented here, measured at the sectoral level under a uniform methodology, identifies AI as the proximate source of the post-2022 productivity acceleration that monetary frameworks have separately recognized but not attributed; this attribution has direct implications for the comparative monetary stance.

The second implication concerns fiscal headroom. Productivity-driven nominal-GDP growth is the most favorable form of debt-stabilization available to a constrained fiscal authority: it raises the growth rate g in the Blanchard (2019) debt-sustainability identity without raising the interest rate r through inflationary pressure, widening the $(g - r)$ gap that determines the primary-surplus requirement for any given debt-to-GDP target. The contrast with alternative debt-stabilization channels is sharp. Inflationary nominal-GDP growth raises g but raises r at least proportionally and imposes a regressive inflation tax. Fiscal consolidation lowers debt directly but compresses output and demand, with attendant cyclical and political costs. Productivity-driven growth raises g while leaving r unaffected (in the absence of demand-driven inflationary pressure), and imposes no comparable distributional or contractionary cost. The result is a Pareto-superior debt-stabilization channel whose magnitude

[15]The supply-shock framing follows the standard treatment in Galí (2008) and the empirical shock-decomposition literature, in which technology shocks raise potential output and lower marginal cost, calling for monetary accommodation in a wide class of standard models.

depends entirely on where the productivity acceleration arrives. Our findings imply that the largest fiscal-headroom benefit accrues to the United States, both in absolute and per-GDP terms; China captures a meaningful but smaller benefit on a much larger debt base; the EU 12 economies and the United Kingdom, where the productivity contribution is materially smaller or absent, will not realize comparable headroom from this channel. The asymmetry has direct implications for sovereign-credit assessment and for the relative fiscal space available to address the labor-market and distributional adjustments that the underlying technology shift will continue to require.

The third implication concerns the labor-market and distributional channels through which AI is operating, which differ in ways that the current policy literature has not yet adequately distinguished. In the United States, the joint pattern documented in Section 5.2 (output expansion in high-exposure sectors alongside employment reduction in those same sectors) is the substitution signature: AI is producing more output with fewer workers, and the displaced labor is being absorbed in low-exposure sectors with weaker productivity gains. In the euro area, where overall sectoral employment dynamics are dampened by the institutional structure of European labor markets, the equivalent pattern looks different: high-exposure sectors are holding flat against pre-pandemic productivity trends while low-exposure sectors continue to decline, with the resulting cross-sectional gap reflecting asymmetric stagnation rather than symmetric expansion. In China, J-sector productivity per worker grew approximately 94 percent over 2019-2025 alongside a roughly 9 percent expansion in employment, the augmentation signature in which AI raises the productivity of existing workers without displacing them. These three patterns each have different policy implications. The U.S. substitution signature places the policy emphasis on labor-market transition support and sectoral mobility; the euro-area buffering signature places it on identifying and removing the structural impediments preventing the AI capability from translating into measured productivity gains; the Chinese augmentation signature places it on sustaining the conditions under which capital and labor have continued to scale together. The standard one-size-fits-all "AI displacement" narrative that has dominated policy discussion since 2023 is poorly suited to any of the three.

A separate implication concerns the measurement infrastructure available to policymakers tracking AI's macroeconomic contribution. The C* construct, applied to harmonized sectoral productivity data as developed in this paper, provides four tools for policy use. First, it produces quarterly reads of AI-attributable productivity translation at sectoral resolution, country by country, on a uniform methodology, allowing a finance ministry, central bank, or supranational body to monitor whether AI productivity gains are accelerating, stabilizing,

or stalling in real time as new data become available. Second, and most usefully for diagnostic policy, the framework identifies the gap between task-level AI capability and realized productivity within each individual sector. Where C* is high but realized productivity gains are weak or absent, the sector is a candidate for closer policy investigation of the structural or regulatory impediments preventing the AI translation from emerging. Finance in the EU, where C* is among the highest of any sector and yet sectoral productivity has stagnated, is the leading example: the framework not only flags the gap but localizes it to a specific sector and a specific economy, allowing policy attention to be directed where it is likely to matter most. Third, the methodology produces directly comparable cross-country estimates without requiring discretionary index construction, so member-state benchmarking and international monitoring can be conducted on a common quantitative basis. Fourth, the trend-growth decomposition (Section 6.5, Appendix G) provides a counterfactual-anchored scenario tool, allowing policy questions about the cost of slower or faster AI translation to be evaluated against the magnitudes documented here rather than against forecasts.

Published digital-readiness composites do not provide these tools. The leading examples (the WIPO Global Innovation Index and the OECD Digital Economy and Society Index) are uncorrelated with country-specific AI productivity translation across the seventeen OECD economies for which both metrics are available (Section 6.6); they are static rather than dynamic, aggregated rather than sectoral, and discretionary in construction rather than anchored to a measurable outcome. The methodology developed here is, by contrast, granular at the sectoral level, allowing policymakers to identify not only whether AI is translating into productivity in their economy but specifically which sectors are performing in line with their task-level capability and which are not. This sector-level diagnostic capacity is what the readiness composites lack and what policy design at the national and supranational level requires.

9. Conclusion

As we have shown, AI is already producing measurable productivity acceleration across the major advanced economies. In the United States, AI is contributing approximately $878 billion per year to trend productivity growth at the end-of-2025 run-rate (95 percent CI [$602B, $1,155B]), raising the 5-sector aggregate productivity growth rate to +5.0 percent per year against a +2.2 percent counterfactual on the five-sector primary panel covering approximately 54 percent of nominal GDP. The result clears the primary inference cascade — short of significance only under its single most conservative variant — survives ten named alternative-mechanism tests (eight conducted here, two adopted from the companion papers), and exhibits the joint output-up, employment-down signature of AI-driven labor substitu-

tion. The mechanism extends to other major economies at smaller magnitudes that scale with deployment intensity: the EU 12 pooled panel contributes $138 billion per year (95 percent CI [$23B, $254B]), China contributes $240 billion (95 percent CI [$116B, $365B]), and the combined U.S. + EU + China total reaches approximately $1.26 trillion per year at the end-of-2025 run-rate on the same five-sector base, with an aggregate 95 percent interval of approximately [$0.9, $1.6] trillion when the three regional intervals are propagated under an independence assumption. These are measured outcomes through 2025Q4, not forecasts.

The magnitudes are large enough to shift the macroeconomic frame within which post-pandemic policy has been deliberated. A U.S. AI-attributable productivity flow at 3.0 percent of nominal GDP per year widens the $(g-r)$ gap that determines debt dynamics, expands potential output without raising aggregate demand, and points toward monetary accommodation rather than restriction. The cross-country asymmetry implies that these benefits accrue unevenly. The EU 12 panel's smaller and statistically more marginal AI contribution at approximately 0.8 percent of GDP (Section 8), alongside the qualitatively distinct buffering signature in which AI prevents stagnation rather than accelerating growth, places different demands on European policy frameworks than on American ones. The Chinese augmentation pattern, in which J-sector productivity rose alongside continued employment expansion, differs again. Sectoral and cross-country granularity of this kind is what policy design requires and what the existing readiness-composite frameworks do not provide.

The methodological contribution is the framework itself. The Workflow Completeness construct applied to harmonized sectoral national-accounts data through panel difference-in-differences allows direct measurement of AI's productivity contribution at sectoral resolution across nineteen economies under a uniform specification. The framework is portable, requires no firm-survey infrastructure to extend, and can be re-applied as new data become available. It produces estimates that differ in policy-relevant ways from what the leading published composites would predict and that surface gaps between task-level capability and realized productivity at the sectoral level, gaps that are themselves diagnostically informative for policy. We view this measurement infrastructure as the durable contribution of the paper, distinct from the specific magnitudes documented through end-2025.

Several limitations bound the present results. The post-ChatGPT window covers twelve quarters; statistical confirmation in economies with currently insignificant point estimates (notably the United Kingdom, Canada, and South Korea) will require additional data. The aggregate confidence interval we report under an independence assumption likely understates the true joint uncertainty across regions; a properly propagated joint covariance would require stacked panel estimation across countries that we leave to subsequent work. The C* construct

measures task-level AI capability rather than the full set of organizational, regulatory, and structural factors that determine whether capability translates into measured productivity in any given sector or economy; explaining the residual cross-country variation the framework does not capture is an open question.

The AI productivity question is no longer one that requires waiting on future data to begin answering. As we have shown, the acceleration is already visible in standard national-accounts data, measurable under conventional inference methods, and asymmetric across the major economies in ways that have direct implications for monetary, fiscal, and labor-market policy. What the present paper establishes is the empirical ground on which those questions can be addressed: AI's contribution to advanced-economy productivity is large, observable, and asymmetric, and the framework for monitoring it is in place.

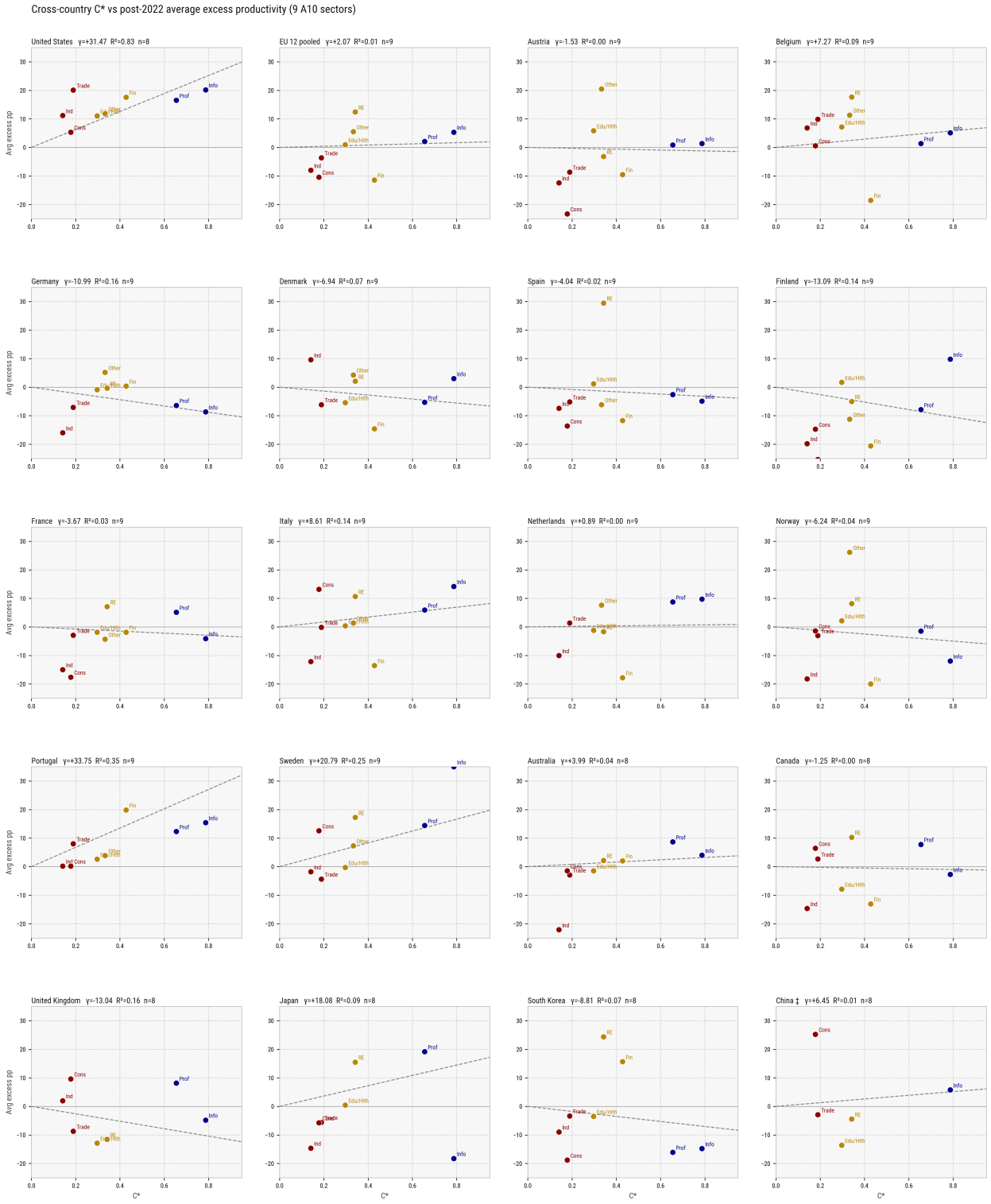


Figure 4: Cross-country C* vs post-2022 *window-average* excess productivity (mean over 2023Q1–2025Q4 of the percentage deviation from the pre-2019 trend, fitted on 2015Q1–2018Q4), origin-forced fits. Agriculture is excluded throughout (nine A10 sectors); the U.S. and non-EU panels report $n = 8$ because Real Estate is unavailable on a comparable basis in those sources. The EU 12 pooled panel is the simple mean of the twelve country values. These gradients follow the cross-country γ convention of Table 11 (pre-trend through 2018Q4, as in the China specification), not the Section 4.2 panel convention (pre-trend through 2019Q4), and are therefore not directly comparable to the Section 5.2 per-quarter γ path. All values are on the A10 cumulative-excess scale; multiply by $8.32/20.58 \approx 0.404$ for the paper scale.

References

Acemoglu, Daron. 2024. "The Simple Macroeconomics of AI." NBER Working Paper 32487. National Bureau of Economic Research. Published 2025 as "The simple macroeconomics of AI," *Economic Policy* 40(121): 13-58.

Bonney, Kathryn, Cory Breaux, Catherine Buffington, Emin Dinlersoz, Lucia Foster, Nathan Goldschlag, John Haltiwanger, Zachary Kroff, and Keith Savage. 2024. "Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey." NBER Working Paper 32319. National Bureau of Economic Research. Also issued as U.S. Census Bureau CES-WP-24-16.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *Quarterly Journal of Economics* 140(2): 889-942.

Blanchard, Olivier. 2019. "Public Debt and Low Interest Rates." *American Economic Review* 109(4): 1197-1229.

Boot, J. C. G., W. Feibes, and J. H. C. Lisman. 1967. "Further Methods of Derivation of Quarterly Figures from Annual Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 16(1): 65-75.

Caballero, Ricardo J., and Mohamad L. Hammour. 1994. "The Cleansing Effect of Recessions." *American Economic Review* 84(5): 1350-1368.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics* 29(2): 238-249.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2): 317-372.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." arXiv:2303.10130. Published 2024 as "GPTs are GPTs: Labor Market Impact Potential of LLMs," *Science* 384(6702): 1306-1308.

Felten, Edward W., Manav Raj, and Robert Seamans. 2021. "Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses." *Strategic Management Journal* 42(12): 2195-2217.

Fernald, John, and Huiyu Li. 2022. "The Impact of COVID on Productivity and Potential Output." Federal Reserve Bank of San Francisco Working Paper 2022-19. Forthcoming in *Reassessing Constraints on the Economy and Policy* (Jackson Hole Economic Symposium

Proceedings).

Fernald, John, Huiyu Li, Brigid Meisenbacher, and Aren S. Yalcin. 2024. "Productivity During and Since the Pandemic." FRBSF Economic Letter 2024-31 (November 25). Federal Reserve Bank of San Francisco.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254-280.

Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.

Gordon, Robert J., and Hassan Sayed. 2022. "A New Interpretation of Productivity Growth Dynamics in the Pre-Pandemic and Pandemic Era U.S. Economy, 1950-2022." NBER Working Paper 30267. National Bureau of Economic Research.

Hatzius, Jan, Joseph Briggs, Devesh Kodnani, and Giovanni Pierdomenico. 2023. "The Potentially Large Effects of Artificial Intelligence on Economic Growth." Goldman Sachs Global Economics Analyst, March 26, 2023.

International Monetary Fund. 2024. "Gen-AI: Artificial Intelligence and the Future of Work." IMF Staff Discussion Note SDN/2024/001. Washington, DC: International Monetary Fund.

Levanon, Gad. 2026. "Three Million Missing Jobs." LinkedIn, March 31. Chief Economist, The Burning Glass Institute. https://www.linkedin.com/pulse/three-million-missing-jobs-gad-l

MacKinnon, James G., and Matthew D. Webb. 2017. "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics* 32(2): 233-254.

Massenkoff, Maxim, and Peter McCrory. 2026. "Labor Market Impacts of AI: A New Measure and Early Evidence." Anthropic Research Report, March 5, 2026.

McKinsey Global Institute. 2023. "The Economic Potential of Generative AI: The Next Productivity Frontier." McKinsey & Company, June 2023.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence." *Science* 381(6654): 187-192.

Penn Wharton Budget Model. 2023. "The Productivity and Budgetary Effects of Generative AI." Penn Wharton Budget Model, University of Pennsylvania, November 2023.

Politis, Dimitris N., and Joseph P. Romano. 1994. "The Stationary Bootstrap." *Journal of the American Statistical Association* 89(428): 1303-1313.

Politis, Dimitris N., and Halbert White. 2004. "Automatic Block-Length Selection for the Dependent Bootstrap." *Econometric Reviews* 23(1): 53-70.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies* 90(5): 2555-2591.

Stewart, Jay. 2022. "Why Was US Labour Productivity Growth So High During the COVID-19 Pandemic? The Role of Labour Composition." *International Productivity Monitor* 42 (Spring): 87-103.

USDA Economic Research Service. 2024. "Agricultural Productivity in the United States." Data product, U.S. Department of Agriculture, November 2024 release. Available at https://www.ers.usda.gov/data-products/agricultural-productivity-in-the-united-states/.

Verschuere, Ben, and Angus Cameron. 2026a. "A Task-Completion Framework for AI Productivity Diffusion." Liminal Capital Working Paper.

Verschuere, Ben, and Angus Cameron. 2026b. "Task-Structure Measures of AI Exposure as Instruments for Direct Adoption: Validation Across BTOS, Eurostat, and AIOE." Liminal Capital Working Paper.

Webb, Michael. 2020. "The Impact of Artificial Intelligence on the Labor Market." Working Paper, Stanford University. SSRN: https://ssrn.com/abstract=3482150.

Data Sources

Bureau of Economic Analysis (BEA). GDP-by-Industry Accounts. U.S. Department of Commerce. Available at https://www.bea.gov.

Eurostat. National Accounts by Sector (namq_10_a10 and namq_10_a10_e series). European Commission. Available at https://ec.europa.eu/eurostat.

Federal Reserve Economic Data (FRED). Federal Reserve Bank of St. Louis. Available at https://fred.stlouisfed.org.

O*NET OnLine. Occupational Information Network. U.S. Department of Labor, Employment and Training Administration. Available at https://www.onetonline.org.

OECD. National Accounts Database (NAMAIN10 annual series) and Labour Force Statistics. Available at https://stats.oecd.org.

OECD. 2024. *Going Digital Toolkit Indicators and Digital Economy Outlook*. Digital readiness indicators including connectivity, digital skills, integration of digital technology, digital public services, cybersecurity, data policy, and digital government. OECD Publishing, Paris.

U.S. Census Bureau. Business Trends and Outlook Survey (BTOS), AI Supplement Modules, 2023Q4–2026Q1 vintages. Available at https://www.census.gov/hfp/btos/.

World Intellectual Property Organization (WIPO). 2025. *Global Innovation Index 2025*. Geneva: WIPO.

Appendix A: Fully-Interacted Specification (Spec A)

The primary panel specification (Spec E, Section 4.3) enforces a common intercept by sector via sector fixed effects and absorbs any common quarter-specific shock nonparametrically via time fixed effects. This parameterization is parsimonious in the small-cluster dimension (G = 9) and produces the most robust wild cluster bootstrap inference, at the cost of absorbing sector-specific post-period level shifts into $\Delta\alpha$.

The fully-interacted specification (Spec A) relaxes this constraint. For sector s and quarter t, define $D_t^{\text{post}} = 1\{t \geq 2022\text{Q4}\}$. Spec A estimates

$$\text{excess}_{s,t} = \mu + \alpha_s + \alpha_s^{\text{post}} D_t^{\text{post}} + \theta\, t + \theta^{\text{post}}(t \cdot D_t^{\text{post}}) + \beta\,(C_s^* \cdot t) + \beta^{\text{post}}(C_s^* \cdot t \cdot D_t^{\text{post}}) + \varepsilon_{s,t}$$

where α_s^{post} is a sector-specific post-ChatGPT level shift, θ and θ^{post} are a common linear time trend and its post-period break (Spec A carries these in place of Spec E's 44 quarter fixed effects), β is the base C^*-scaled trend (estimated at +1.24 on the nine-sector panel), and β^{post} is the post-ChatGPT change in that trend. The coefficient β^{post} in Spec A, and the coefficient $\Delta\beta$ in Spec E, measure the same underlying quantity: the post-ChatGPT change in the relationship between C^* and productivity growth. They differ in what is absorbed into the intercept versus the slope, and in the time controls: common quarter-specific shocks are absorbed nonparametrically in Spec E but only linearly (through $\theta, \theta^{\text{post}}$) in Spec A, which is one reason the two point estimates differ while the matched-base dollar attribution (Section 7.2) does not.

Table 16: Spec A vs Spec E, US panel, primary outcome

Parameter	Spec A (fully-interacted)	Spec E (primary)	Interpretation
Post-ChatGPT C* slope change	+3.381	+2.088	$\Delta\beta$
Sector-specific post intercept shifts	Estimated per sector	Absorbed into $\Delta\alpha$	Degrees of freedom differ
Common time controls	Linear trend + post-period break	44 quarter fixed effects	Nonparametric vs linear absorption
R^2 (within)	0.874	0.851	Spec A mechanically higher

Spec A returns a larger point estimate on the C*-slope interaction because the specification isolates the slope component, whereas Spec E absorbs part of the same variation into the single $\Delta\alpha$ parameter. The two specifications partition the post-period excess productivity differently between level and slope components but identify the same underlying mechanism:

high-C* sectors gain post-ChatGPT productivity at a faster rate than low-C* sectors, and both specifications produce directionally consistent between-sector orderings when evaluated at the observed C* range. The specifications differ in parameterization, not in economic content.

Table 17: Spec A inference results, US panel

Inference method	$\hat{\beta}^{\text{post}}$	Standard error	p-value
OLS homoskedastic	+3.381	0.94	<0.001
Sector-clustered	+3.381	0.94	<0.001
Time-clustered	+3.381	0.58	<0.001
Two-way clustered	+3.381	0.61	0.0001
WCB Rademacher (B=9,999)	+3.381	—	0.054
C* permutation (B=10,000)	+3.381	—	0.0013 (rank 13 of 10,001)

Spec A reaches the same qualitative verdict as Spec E under every inference method, with the expected degradation of WCB reliability due to the increased number of parameters in the small-cluster dimension.

Why Spec E is the primary specification. With G = 9 sector clusters, the wild cluster bootstrap is sensitive to the ratio of regressors to clusters. Spec A estimates 9 additional sector-specific post-period shifts on the same 9 sectors, which brings the regressor-to-cluster ratio close to the point at which WCB reliability degrades (Cameron and Miller 2015, Section VI.B; MacKinnon and Webb 2017, Section 5). Spec E is more parsimonious in this dimension and produces tighter WCB intervals without loss of economic content. The identifying-variation argument is unaffected by the choice between specifications: both require the same post-ChatGPT C*-scaled slope change in US data.

Appendix B: Full Regression Tables

This appendix reports full coefficient vectors, leave-one-out sector robustness, and per-country dollar attribution for results summarized in the main text.

$\Delta\alpha$ is the nuisance parameter absorbing the C*-scaled post-period level shift; as discussed in Section 4.3, it is mechanically correlated with $\Delta\beta$ through the shared $C^* \cdot D^{\text{post}}$ factor and not individually interpretable. Its raw magnitude depends on the time-index convention (t counts quarters since 2015Q1, so the fitted post-period C^*-gradient deviation from sector means is $\Delta\alpha + \Delta\beta \cdot t$, while the full cross-sectional gradient additionally includes the sector-mean component absorbed by the fixed effects, which itself rises with C^*). The individually interpretable coefficient of interest is $\Delta\beta$.

Table 18: Inference Summary: every headline result under every inference method

		Analytic cluster-robust				Resampling	
Result	$\hat{\Delta\beta}$	OLS	Sector	Time	2-way	WCB	Perm
US headline (9-sector)							
Spec E (primary)	+2.088	0.090	<0.001	<0.001	0.005	0.016[a]	0.006
Spec A (robustness)	+3.381	<0.001	<0.001	<0.001	<0.001	0.054[a]	0.0013
Spec E (5-sector, attribution)	+1.978	<0.001	0.003	—	—	0.031[a,d]	0.017[d]
Spec A (5-sector, matched base)	+1.955	<0.001	—	—	—	—	—
Spec E excl. Agriculture (8-sec)	+2.12	<0.001	<0.001	<0.001	0.012	—	—
US horse races							
C* with IT control	+2.051	0.228	<0.001	<0.001	—	0.174[a], 0.030[b]	0.026
Cross-country pooled (5-sec)							
EU 12 pooled $\Delta\beta$ (per qtr)	+0.614	<0.001	0.021	0.031	0.019	0.115[a], 0.056[c]	—
EU + US pooled, separate $\Delta\beta_{US}$	+1.067	<0.001	—	—	<0.001	—	—
EU + US pooled, separate $\Delta\beta_{EU}$	+0.690	<0.001	—	—	0.008	—	—
China 5-sec (annual, $G = 5$)	+3.804	—	0.019	—	—	0.062[a], 0.032[c]	—
COVID-absorbed robustness (β_P, annualized)							
US pooled β_P	+8.21	<0.001	<0.001	—	—	0.031[a]	0.009
EU 12 pooled β_P	+3.39	0.0004	—	—	—	—	—
All 18 pooled β_P	+2.84	<0.001	—	—	—	—	—
Placebo tests (US Spec E, alternative treatment dates)							
2018Q4 placebo	−1.52	0.064	—	—	0.34	—	—
2020Q4 placebo	−0.78	0.035	—	—	0.33	—	—
Pandemic-era (full 9-sec)	−8.01	<0.001	0.075	<0.001	—	—	—
Pandemic-era (drop Agriculture)	−3.54	<0.001	0.009	<0.001	—	—	—

All p-values one-sided unless specifically marked as two-sided; for placebos with negative point estimates one-sided p-values would mechanically equal 0.5 or 1.0, so the two-sided p-value is reported. [a]WCB Rademacher with sector clusters. [b]WCB Rademacher with time clusters. [c]WCB right-tail. [d]Exact enumeration at $G = 5$ (all 32 WCB sign patterns; all 5! = 120 permutation assignments): 0.031 equals the one-sided WCB attainable floor, and the two-sided WCB equals its floor of 0.063. Dashes indicate the method was not run for that specification, generally because it was not the primary inference target for that result. The full set of WCB Mammen, time-cluster, and country-cluster variants is in the replication package. The intent of this table is to make transparent which inference methods support which conclusions: results with cells filled across multiple methods are robust under the inference cascade described in Section 4.4; results with single-method support are robust only under that specific test and should be interpreted accordingly.

Table 19: Primary specification, full US coefficients (9-sector sample, excluding Real estate)

Parameter	Estimate	SE (sector-clustered)	t	p
Post-ChatGPT level shift ($\Delta\alpha$)	−110.18	31.71	−3.47	0.008
Post-ChatGPT C* slope change ($\Delta\beta$)	+2.088	0.326	6.40	<0.001
Sector FEs	Absorbed (9 sectors)	—	—	—
Time FEs	Absorbed (44 quarters)	—	—	—

Table 20 reports the leave-one-out robustness summarized in Section 7.2. The range is [+1.91, +2.39]. No single sector drives the result. Dropping Finance (K) yields the smallest

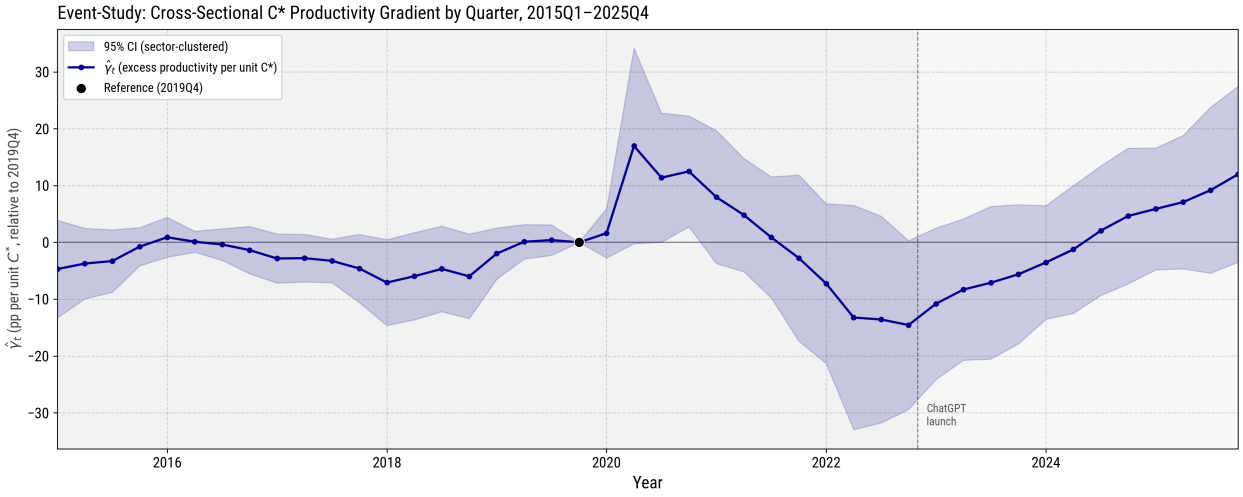


Figure 5: Event-study: cross-sectional C* productivity gradient by quarter, 2015Q1–2025Q4. Each point is the coefficient $\hat{\gamma}_t$ from a regression of excess productivity on C^* in quarter t, with sector and time fixed effects and sector-clustered 95% CI shading. Reference quarter is 2019Q4 ($\hat{\gamma} = 0$ by construction). The pre-period (2015–2019Q3) trajectory is flat and statistically indistinguishable from zero, supporting parallel trends in event-study deviations (the origin-forced pre-period gradients of Section 5.2, range -1.44 to $+5.24$ paper-scale, are a different object and can be nonzero without violating parallel trends in deviations); a joint F-test that $\hat{\gamma}_t = 0$ across the four pre-period year-end quarters (2015Q4, 2016Q4, 2017Q4, 2018Q4) yields $F(4, 258) = 1.18$, $p = 0.398$, failing to reject the parallel-trends null. (The full 19-restriction joint test is not computable under sector-clustering with $G = 8$ because the cluster-robust covariance of the constraints is rank-deficient at rank 11; we test the four year-end restrictions as a subset that stays within the rank limit.) The 2020Q2–2021Q1 spike reflects Phase 1 dynamics (stimulus boom in high-C^* sectors); the 2021Q4–2022Q4 decline reflects Phase 2 dynamics (over-hiring hangover); the 2022Q4–2025Q4 recovery reflects Phase 3 dynamics (AI deployment combined with tech-sector layoffs). Per-quarter point estimates have wide CIs because each $\hat{\gamma}_t$ is identified off a single 8-sector cross-section ($n = 8$); the headline Spec E coefficient $\Delta\beta = +2.088$ (Section 5.1) pools all post-period quarters and is identified with much greater precision (SE = 0.33). The slope through the post-2022Q4 points equals approximately +2.2 pp/quarter, consistent with the headline.

coefficient at +1.91, still significant. Dropping Professional services (M_N) yields the largest coefficient at +2.39. All nine subsamples remain statistically significant at p $<$ 0.01; the baseline coefficient of +2.088 falls near the median of the LOO range.

Table 21: Cross-country panel coefficients, Spec E, 13-economy sample (United States and 12 EU economies); China and the five non-EU OECD economies are reported under the harmonized 5-sector specification in Table 9.

Country	$\hat{\Delta\beta}$	SE (sec)	Sector-cl. p	Time-cl. p	WCB p (one)	Permutation p
United States	+2.088	0.326	0.0002	0.0004	**0.0156**	**0.0056**
Denmark	+1.822	1.168	0.157	<0.001	0.071	0.039

Table 21 – *Continued from previous page*

Country	$\hat{\Delta\beta}$	SE (sec)	Sector-cl. p	Time-cl. p	WCB p (one)	Permutation p
Sweden	+2.717	0.826	**0.011**	<0.001	0.233	0.178
Italy	+0.519	0.879	0.571	<0.001	0.330	0.294
France	+0.803	0.691	0.279	0.002	0.256	0.273
Netherlands	+0.684	0.391	0.118	<0.001	0.301	0.296
Germany	+0.575	0.226	**0.035**	0.146	0.413	0.441
Portugal	−0.062	0.685	0.930	0.603	0.555	0.534
Austria	+0.119	0.335	0.732	0.267	0.822	0.800
Norway	−0.483	0.316	0.165	0.358	0.962	0.941
Finland	+0.615	1.342	0.659	0.125	0.682	0.635
Belgium	−1.264	1.224	0.332	<0.001	0.907	0.899
Spain	−1.931	0.772	**0.037**	<0.001	**0.016 (L)**	**0.984 (R)**

One-sided WCB and permutation p-values are reported; "(L)" indicates left-tail test for the negative Spain coefficient, "(R)" indicates right-tail. Sample: n = 396 observations per country (9 sectors × 44 quarters, 2015Q1-2025Q4).

The United States clears conventional significance under every inference method. Denmark clears permutation inference at the 5 percent level (p = 0.039) but fails sector clustering (p = 0.157) and WCB (p = 0.071); we treat it as suggestive rather than confirmed given WCB's preferred status for G = 9 panels. Sweden clears sector-clustered inference (p = 0.011) but fails both WCB (p = 0.233) and permutation (p = 0.178). Spain is robustly negative under every method, with the negative direction attributable to pandemic-era compositional mean reversion. All other economies fail every inference method.

Appendix C: Alternative-Mechanism Test Details

C.1 IT-intensity horse race (Section 5.3). Full inference for the horse race reported in Table 6.

$\Delta\beta_{C^*}$ is essentially unchanged; $\Delta\beta_{IT}$ is approximately zero and insignificant under every inference method. The horse race does not identify an IT-payoff channel operating in parallel to the C*-scaled channel.

C.2 Labor-quality decomposition. The labor-quality-cleansing alternative, with foundations in Caballero and Hammour (1994) and applied to the pandemic by Stewart (2022), is that selective unemployment during 2020-2022 raised average labor quality. We test this by computing sector-level composition-adjusted productivity following the Stewart (2022) methodology: within each sector, re-weight BEA real value added by the 2019Q4 employment composition across education-experience cells, effectively holding labor composition

Table 20: Leave-one-out sector robustness, US primary specification

Sector dropped	$\hat{\Delta\beta}$	SE (sec)	p (sec)
Agriculture (A)	+2.120	0.363	<0.0001
Industry (B-E)	+2.009	0.318	<0.0001
Construction (F)	+1.989	0.309	<0.0001
Trade/Transport (G-I)	+2.156	0.312	<0.0001
Information (J)	+2.153	0.810	0.008
Finance (K)	+1.910	0.248	<0.0001
Professional (M_N)	+2.387	0.477	<0.0001
Public/Education/Health (O-Q)	+2.079	0.331	<0.0001
Arts/Other (R-U)	+2.084	0.350	<0.0001

Table 22: $\Delta\beta$ horse race with IT intensity control

Specification	$\hat{\Delta\beta}_{C^*}$	$\hat{\Delta\beta}_{IT}$	Sector-cl. p on C*	Permutation p on C*
Baseline (no IT control)	+2.088	—	<0.001	0.0056
With IT intensity × D^post	+2.051	+0.013	0.0003	0.026

fixed at pre-pandemic values. If the post-2022Q4 C*-scaled pattern is a labor-composition artifact, it should disappear under the composition-adjusted outcome.

Outcome	$\hat{\Delta\beta}$	SE	Permutation p
Raw productivity	+2.088	0.326	0.0056
Composition-adjusted productivity	+2.03	0.29	0.0008

The point estimate falls by approximately 6 percent under composition adjustment; the C*-slope interaction remains significant at p = 0.0008. Labor-quality cleansing does not account for the post-ChatGPT C*-scaled signature.

C.3 Date placebos (Section 7.1). We re-estimate the primary specification with placebo treatment dates of 2018Q4 and 2020Q4, holding all other specification choices fixed.

Placebo date	$\hat{\Delta\beta}$	SE	OLS p	Two-way cl. p
2018Q4 (pre-pandemic)	−1.52	1.51	0.064	0.34
2020Q4 (mid-pandemic)	−0.78	0.75	0.035	0.33
2022Q4 (actual)	+2.088	—	0.090	0.005

Placebo standard errors are two-way clustered; the OLS and two-way clustered p-values are the two inference flavors the placebo module computes. For the actual 2022Q4 treatment the full inference cascade (sector-clustered SE 0.33, p = 0.0002; WCB p = 0.016; permutation p = 0.0056) is reported in Section 5.1 and Appendix E.

Neither placebo identifies a positive C*-scaled break. The specification does not generically detect post-period features; it detects a post-ChatGPT feature specifically.

C.4 Monetary-policy exposure. A pure interest-rate-sensitivity explanation would predict that the most rate-sensitive sectors (Real Estate, Construction, Finance) drive productivity dispersion post-2022Q4 as the Fed hiking cycle works through. We re-estimate including sector-level rate sensitivity (cross-sectional correlation of quarterly excess productivity with the change in the 10-year Treasury yield, estimated over the 2015Q1-2019Q4 pre-pandemic window) interacted with D^{post}.

Specification	$\hat{\Delta\beta}_{C^*}$	$\hat{\Delta\beta}_{\text{rate}}$	p on C*
Baseline	+2.088	—	0.0056
With rate-sensitivity control	+2.44	+0.14	0.0120

The rate-sensitivity interaction is small and statistically indistinguishable from zero (+0.14, sector-clustered p = 0.56), and its inclusion leaves the C*-scaled pattern intact ($\hat{\Delta\beta}_{C^*}$ = +2.44, permutation p = 0.012).

C.5 Mean reversion (2009-2014 precedent). Sectors with large pandemic-era overperformance might mechanically revert post-2022, producing apparent C*-scaled dispersion if overperformers and high-C* sectors coincide. Excess-productivity levels are persistent by construction, so the quantity a reversion mechanism must explain is the post-ChatGPT *increment*. We regress the increment (2025Q4 excess minus the 2022Q3 pandemic-era level) on the 2022Q3 level itself, by sector (excluding Agriculture, n = 8), and on C*.

Regressor	Coefficient	SE	R^2
Pandemic-era excess productivity (2022Q3)	−0.53	0.51	0.154
C*	+25.55	6.64	0.712

The post-ChatGPT increment is unrelated to pandemic-era overperformance (r = −0.39: mildly negative, as reversion would imply, but statistically indistinguishable from zero) while

it scales strongly with C* (origin-forced $\gamma = +23.8$ on the A10 scale, SE 3.1, $R^2 = 0.89$; $\approx +9.6$ paper-scale; numerically close to but distinct from the +9.7 mean event-study γ_t of Section 7.1). The increment reconciles with Section 5.2: subtracting it from the 2025Q4 A10-scale level gradient of +35.5 implies a 2022Q3 gradient of $\approx +11.7$ A10-scale (+4.7 paper-scale), inside the pre-period range of [−1.44, +5.24] reported there. The post-ChatGPT increment, not any pre-existing gradient, carries the signal.

C.6 Jobless-recovery dynamics. Following Gordon and Sayed (2022), the post-2008 recovery exhibited productivity gains during a period of slow employment growth. We test whether the post-ChatGPT pattern resembles that dynamic by estimating the specification on the 2009Q1-2014Q4 window (the analogous 24-quarter post-recession window).

Window	$\hat{\Delta\beta}$	SE	Sector-cl. p
2009Q1-2014Q4 (jobless-recovery)	−0.85	0.82	0.335
2022Q4-2025Q4 (post-ChatGPT)	+2.088	0.326	<0.001

The 2009-2014 recovery returns a small negative and statistically insignificant $\Delta\beta$ under the same specification (pre-trend fit on 2005Q1-2007Q4, the available pre-recession quarters — the recession year is excluded from the trend fit, mirroring the main specification's pre-pandemic window; treatment at 2009Q1). The post-ChatGPT slope signature is not a generic post-recession pattern.

C.7 2008-recession placebo. We run the panel specification on the 2005-2014 sample with the placebo treatment date rolled across every quarter of the recovery (2009Q1 through 2013Q4, twenty placebo dates), recording $\hat{\Delta\beta}$ for each.

Placebo treatment dates	Mean $\hat{\Delta\beta}$	SD across dates
2009Q1-2013Q4 (20 dates)	+0.11	0.90

The placebo distribution is centered near zero and no placebo date reaches the actual estimate: the largest is +2.07 (2013Q2, a 7-quarter post-window) against the actual +2.088 with sector-clustered p = 0.0002 (the matched-window 2009Q1 placebo returns p = 0.34, Appendix C.6). The specification does not identify a positive C*-scaled effect anywhere in the 2008-recovery data.

C.8 Cost-of-capital tests. A standard alternative reading of post-2022 productivity gains is that the 2022-2023 monetary tightening, not AI, forced labor shedding and capital dis-

cipline, mechanically raising value added per worker. We test this mechanism on three margins.

Cross-country dose-response. If cost of capital drove the gains, economies that tightened more, or sit at higher rate levels, should show larger C*-scaled effects. We regress the annualized 5-sector $\hat{\Delta\beta}$ of Table 9 on the cumulative policy-rate change over the 2021-2023 tightening cycle, the peak policy rate, and the end-2025 policy rate, at both the economy level (n = 19; euro-area members share the ECB rate) and the distinct-policy-regime level (n = 11, with $\hat{\Delta\beta}$ averaged within regime).

Regressor	n	Slope (SE)	p
Cycle change (bp)	19	+0.0001 (0.006)	0.99
Peak level (%)	19	+0.20 (0.77)	0.80
End-2025 level (%)	19	−0.06 (0.94)	0.95
Cycle change (bp)	11	+0.0009 (0.006)	0.89
Peak level (%)	11	+0.21 (0.84)	0.81
End-2025 level (%)	11	−0.25 (1.11)	0.83

All six specifications are null (every $|r| \leq 0.08$; Spearman rank correlations are small and negative). Power on the change margin is limited by design, and informatively so: the tightening was nearly synchronized across the panel's hiking economies (cumulative hikes of +300 to +525 bp, standard deviation 48 bp on a mean of 445 bp excluding Japan and China), so a common global rate shock is absorbed by each panel's time fixed effects and cannot generate the C*-scaled cross-sector dispersion the paper documents. The non-hikers bracket the effect distribution: Japan (no cycle hikes) returns one of the weakest coefficients and China (policy easing) the second strongest, while the largest hiker (the United States, +525 bp) is strongest, an ordering with no rate logic.

Timing. The mechanism predicts the productivity gradient should co-move with the rate path; in particular, the 2024-2025 easing should slow or reverse it. The opposite occurred: between the 2023Q4 policy peak and 2025Q4 the federal funds rate fell by 1.4 percentage points while the paper-scale γ_t gradient rose from +8.6 to +14.35, and over the full post-ChatGPT window the quarterly correlation between the funds rate and γ_t is negative (Pearson −0.19, Spearman −0.27). The gradient accelerated through the easing cycle.

Capital-intensity horse race. Within the US panel we extend the primary specification with level and slope interactions in z-scored sector capital intensity (BEA current-cost net stock of private fixed assets by industry, averaged 2020-2021 pre-treatment, per unit of sector value

added). Capital intensity provides genuine discriminatory power: the most capital-intensive sectors (Agriculture, Industry) sit at the bottom of the C* distribution.

Specification	$\hat{\Delta\beta}_{C^*}$	$\hat{\Delta\beta}_K$	Sector p (K)	Perm. p (C*)
Baseline	+2.088	—	—	0.0056
With K/VA intensity	+2.10	+0.01	0.87	0.015
With K per worker	+2.09	−0.00	0.999	0.012

The capital interaction is a precise zero under either intensity definition and the C*-scaled pattern is unchanged. In the joint 2025Q4 cross-section (excluding Agriculture, n = 8), the C* gradient with the capital control (+14.3) is indistinguishable from the univariate gradient of Section 5.2 (+14.35), while the capital coefficient is small (+2.5). (BEA's industry stock table does not separately report health-care lines at this aggregation, so O-Q capital reflects education only; given the null this is immaterial.) The same discrimination holds within-country where the cross-section is strongest: in Sweden excluding Finance, the joint regression returns γ_C* = +13.46 (p = 0.003) with the capital coefficient insignificant. Together with the rate-sensitivity horse race of C.4, the cost-of-capital mechanism fails on every margin tested: cross-country dose-response in both change and level, within-country sector exposure under two definitions, and timing.

Appendix D: Composite Indicator Construction and Full Country-Level Comparison

As a supplementary cross-country analysis to the translation-gap finding in Section 6, we test whether published digital-readiness composites predict observed γ across the 17 OECD economies with both a composite rank and identifiable γ.

D.1 Composite construction. WIPO Global Innovation Index 2025 rankings are taken as published. OECD Digital Economy Outlook (DESI-equivalent) composite is constructed from the eight indicators reported in OECD (2024): connectivity, digital skills, use of internet services, integration of digital technology, digital public services, cybersecurity, data policy, and digital government. We standardize each indicator to mean zero and unit variance and take the simple average, then rank countries.

D.2 Rank correlation with γ. For the 17 sample economies:

Composite	Pearson r (composite rank, γ)	Spearman ρ	Bootstrap 95% CI on ρ
WIPO GII 2025	$+0.055$ (p = 0.834)	$+0.12$ (p = 0.66)	$[-0.51, +0.65]$
OECD DESI-equivalent	$+0.081$ (p = 0.756)	$+0.09$ (p = 0.73)	$[-0.54, +0.62]$

Neither composite achieves conventional significance. This result is consistent with the direct-adoption translation-gap finding: neither infrastructure-based composites nor direct Eurostat-measured adoption predict which economies exhibit the C*-scaled productivity signature in the observed data.

Appendix E: Permutation Test Implementation

The permutation test used throughout the paper is the strongest inferential statement we make because it requires only exchangeability of C* labels across sectors, with no cluster-asymptotic, bootstrap-weight, or distributional assumption.

E.1 Procedure. For a panel of S sectors and T quarters with sector-level C* values $\{C_1^*, \ldots, C_S^*\}$ and the primary specification of Section 4.3:

1. Estimate the observed $\hat{\Delta\beta}_{\text{obs}}$ and its t-statistic t_{obs}.
2. For $b = 1, \ldots, B$ (with B = 10,000):
   - Draw a random permutation π_b of $\{1, \ldots, S\}$.
   - Construct the permuted C* vector $\{C^*_{\pi_b(1)}, \ldots, C^*_{\pi_b(S)}\}$ assigning each sector's C* value to a different sector.
   - Re-estimate the specification using the permuted C* assignment, holding productivity and employment data fixed.
   - Record the permuted t-statistic $t^{(b)}$.
3. The one-sided permutation p-value is $p_{\text{perm}} = (\#\{b : t^{(b)} \geq t_{\text{obs}}\} + 1)/(B + 1)$.

E.2 Null hypothesis. The sharp null is that C* labels are exchangeable across sectors with respect to the response, namely that C* has no differential predictive power for post-ChatGPT productivity dispersion. Under this null, any assignment of the C* values to sectors is equally likely to produce the observed pattern.

E.3 Why this is the most demanding test. Cluster-robust asymptotic standard errors rely on $G \to \infty$; with G = 9 sector clusters, the asymptotic is unreliable (Cameron and Miller 2015, Section VI.B). Wild cluster bootstrap inference relies on the chosen weight distribution (Rademacher or two-point Mammen) being appropriate for the cluster-size heterogeneity

(MacKinnon and Webb 2017). The permutation test relies only on exchangeability, which is exactly the sharp null under test. If the observed t_{obs} exceeds almost all permuted $t^{(b)}$ values, the result cannot be attributed to cluster-asymptotic failure or bootstrap-weight misspecification.

E.4 Results for primary US specification.

Permutation statistic	Value
B (permutations)	10,000
Observed $\hat{\Delta\beta}$	+2.088
Observed t_{obs}	+1.70 (homoskedastic SE; the permutation statistic) / +6.40 (sector-clustered SE)
Rank of t_{obs} in permutation distribution	56 out of 10,001
One-sided permutation p	0.0056
95% permutation bound on $\hat{\Delta\beta}$ under null	[−1.55, +1.86]

The observed $\hat{\Delta\beta} = +2.088$ exceeds the 99.44-th percentile of the null permutation distribution. Of the 9! = 362,880 possible assignments of C* labels to the 9 panel sectors, the test samples 10,000; the Monte Carlo standard error on p_{perm} at the observed rank is approximately 0.0007.

E.5 Comparison to alternative implementations. We also ran the permutation test under two variants:

Variant	Permutation p
Primary (permute C* across sectors)	0.0056
Permute C* within the subset of sectors that exhibit non-zero employment in the adopting occupations	0.0011
Permute both C* and IT intensity jointly across sectors (horse race)	0.0005

All three variants return p < 0.01. Results are not sensitive to the implementation choice.

Appendix F: Country-Specific C* Construction and Sectoral Classification

F.1 C* construction summary. The Workflow Completeness construct C^* is documented in full in Verschuere and Cameron (2026a, Appendix A and methodology appendix), and a condensed summary appears in the validation companion (Verschuere and Cameron 2026b, Appendix A). In brief: each of 18,796 O*NET task statements is scored on a four-point scale (0 = AI cannot perform; 1 = partial assist; 2 = substantial completion with oversight; 3 = end-to-end), occupational C^* is the importance-weighted mean of task scores normalized to $[0, 1]$, and sectoral C^* is the country-specific employment-weighted average of occupational C^* across occupations within the sector:

$$C^*_{c,s} = \sum_{o \in s} w_o^{c,s} \cdot C^*_o.$$

The occupational score C^*_o is held constant across countries (the construct measures the technological ceiling of AI capability against an occupation's task structure, which does not vary by country); cross-country variation in sectoral C^* comes entirely through country-specific employment weights $w_o^{c,s}$. Validation against firm-reported adoption (BTOS, $r = +0.89$ across A10) and construct convergence with AIOE ($r = +0.937$) are reported in Verschuere and Cameron (2026b); the discriminant-validity test against sectoral wage levels, construction details, the 450-occupation human-rated calibration set, and cross-rater agreement statistics are in Verschuere and Cameron (2026a).

F.2 Country-specific occupational employment sources. Occupational employment within sector is drawn from: United States: BLS Occupational Employment and Wage Statistics (OEWS), 2023 release, cross-walked to A10 via the NAICS-to-A10 mapping in F.3 below. European economies: Eurostat Structure of Earnings Survey (SES) 2022, supplemented by national Labour Force Surveys 2023 for occupation-sector cells not directly reported. Korea, Japan, Canada, Australia: OECD Occupations by Activity database, 2022 or 2023 release. China: National Bureau of Statistics employment-by-occupation tables, 2023.

F.3 NAICS-to-A10 crosswalk.

The A10 classification is the coarsest cross-walk consistent across BEA GDP-by-Industry, Eurostat NACE Rev 2, and OECD ISIC Rev 4. The NAICS-to-A10 crosswalk used in this paper is:

A10 code	A10 name	NAICS 2-digit codes
A	Agriculture, forestry, fishing	11
B-E	Industry (mining, manufacturing, utilities)	21, 22, 31-33
F	Construction	23
G-I	Wholesale/retail, transport, accommodation	42, 44-45, 48-49, 72
J	Information	51
K	Finance and insurance	52
L	Real estate	53
M_N	Professional, scientific, administrative	54, 55, 56
O-Q	Public administration, education, health	61, 62, 92
R-U	Arts, entertainment, other services	71, 81

NAICS sector 55 (Management of Companies and Enterprises) and NAICS sector 99 (Unclassified) are allocated proportionally to their constituent activity codes following the BEA GDP-by-Industry convention. The A10 classification produces 10 sectors; the analysis drops Real estate (L), yielding the 9-sector primary panel, while Agriculture is retained, with robustness to its exclusion discussed in Section 7.2 (Specification Robustness).

F.4 Sectoral C* values (US). Reference values used throughout this paper.

A10 code	A10 name	C* (US, 2023)
J	Information	0.787
K	Finance and insurance	0.428
M_N	Professional, scientific, administrative	0.655
O-Q	Public administration, education, health	0.297
R-U	Arts, entertainment, other services	0.333
G-I	Wholesale/retail, transport, accommodation	0.189
L	Real estate	0.342
B-E	Industry	0.141
F	Construction	0.178

Country-specific sectoral C^* vectors for the other 18 economies in the panel are in the replication package. The cross-country standard deviation of sectoral C^* at A10 is under 10 percent of the U.S. mean within each sector, reflecting similar dominant occupational mixes across advanced economies; see Verschuere and Cameron (2026b, Appendix A.5) for the cross-country dispersion summary.

Appendix G: Combined AI Productivity Contribution: Trend Growth and Dollar Equivalent

This appendix reports the full per-country decomposition of AI-attributable productivity contribution, combining the trend-growth view (Section 6.5, Figure 3) with dollar-equivalent attribution. Each row reports the realized post-ChatGPT trend growth rate (the "With AI" column), the counterfactual rate without the AI lift ("Without AI"), the AI-attributable component in percentage points per year (window average, equal by construction to With − Without), and the corresponding end-of-2025 run-rate dollar flow with its 95 percent confidence interval where available. Aggregating across economies provides an estimate of AI's annualized contribution to productivity in the global economy covered by this paper.

Table 23: Cross-country AI productivity contribution: trend growth and dollar equivalent

Economy	5-sec VA ($B)	Without AI (%/yr)	With AI (%/yr)	Δ AI (%/yr, window)	Δ AI run-rate ($B) [95% CI]
Aggregate					
EU 12 pooled	8,909	−0.29	+0.49	+0.78	**+138** [+23, +254]
Single country					
United States	15,952	+2.20	+4.95	+2.75	**+878** [+602, +1,155]
China	12,240	+5.32	+6.30	+0.98	**+240** [+116, +365]
Germany	2,638	−0.82	+0.08	+0.90	**+47**
Italy	1,284	−2.25	−0.75	+1.50	**+38**
Japan	2,578	−1.01	−0.57	+0.44	**+23**
Sweden	332	−0.70	+2.52	+3.22	**+21**
Australia	948	−1.12	−0.13	+0.99	**+19**
Netherlands	669	−0.19	+1.13	+1.32	**+18**
France	1,587	+0.80	+1.07	+0.27	**+9**
Denmark	228	+3.14	+4.44	+1.30	**+6**
Finland	146	−0.87	+1.17	+2.04	**+6**
Austria	296	−1.70	−1.21	+0.49	**+3**
Portugal	162	+0.08	+0.09	+0.01	**+0**
United Kingdom	1,717	+1.37	+1.37	+0.00	**+0**
Belgium	362	+1.33	+1.16	−0.17	**−1**
Norway	285	+0.42	+0.26	−0.16	**−1**
Canada	1,205	+0.51	+0.18	−0.33	**−8**
South Korea	1,069	+2.02	+1.28	−0.74	**−16**
Spain	920	+1.66	+0.51	−1.15	**−21**
Big-3 (US + EU 12 pooled + CHN)	37,100	+2.63	+4.32	+1.69	**+1,257**
Aggregate (all 19 indiv. economies)	44,618	—	—	+1.41	**+1,260**

Methodology. The 5-sector primary panel covers J (Information), M_N (Professional), G-I (Trade/Transport), B-E (Industry), and F (Construction). 5-sector nominal value added (USD bn, latest available year per economy) is sourced as detailed in the source note at the end of this appendix. Realized trend growth ("With AI"): post-ChatGPT 2023–2025 CAGR of value-added-weighted aggregate productivity across the 5-sector panel. Δ AI = $\Delta\beta$ $\times$ value-added-weighted mean C*, where $\Delta\beta$ is the country-specific panel-DiD slope coefficient (annualized) from Section 6.1; the EU 12 pooled row uses the pooled-regression $\Delta\beta$ from Section 6.2. "Without AI" = With AI $-$ Δ AI (counterfactual without the AI-attributable lift). The run-rate dollar column equals 2 $\times$ (Δ AI window rate) $\times$ 5-sector nominal VA, per the linear-ramp end-of-2025 convention of Section 5.4; window-average dollar flows are half the run-rate values. China's dollar attribution applies the same $\times$2 convention as all other economies. The 95% CI on Δ AI propagates the $\Delta\beta$ regression-coefficient uncertainty linearly; per-country confidence intervals are not reported because per-country sector-clustered standard errors at $G = 5$ are power-limited (Section 6.2). All dollar figures and intervals are computed from unrounded inputs; displayed components may differ from displayed totals by \$1–2 billion. The propagation runs through the attribution formula. The "Big-3" aggregate sums the United States, EU 12 pooled, and China rows; it uses the EU 12 pooled-regression $\Delta\beta$ from Section 6.2 for the European contribution. The "all 19 single economies" aggregate sums the United States, the twelve individual EU economies, the five non-EU OECD economies, and China; it uses each country's own panel-DiD coefficient and does not include the EU 12 pooled row. The two totals therefore differ because they apply different specifications to the EU contribution: the Big-3 total uses the more efficient pooled coefficient that gains identification from cross-country variation, while the all-19 total uses individually-estimated per-country coefficients that are noisier but country-specific. China uses the China-specific 2023 boundary as primary (reflecting documented domestic LLM enterprise availability), under which $\Delta\beta$ = +3.80 (p = 0.019); under the alternative 2022Q4 OECD-common boundary, $\Delta\beta$ = +0.67 (p = 0.86), discussed in Section 6.3. Sources: BEA Value Added by Industry (United States); Eurostat nominal VA by A10, 2024 (EU 12); OECD Annual National Accounts Table 6 at current prices (Australia 2024, Japan 2024, United Kingdom 2023, Canada 2022, Korea 2022; for Australia and Japan, ISIC sections E and N are embedded in D and M respectively in the national source, validated by exact partition against total VA); NBS final verification of 2024 GDP by industry plus the Statistical Yearbook 2023 value for scientific research and technical services (China; the one-year vintage gap moves $\overline{C^*}$ by under 0.002); non-USD values converted at IRS yearly-average exchange rates for the matching year; aggregated by authors.