ISSUE 50

SPRING 2026

INTERNATIONAL Productivity Monitor

Centre for the Study of Living Standards

THE PRODUCTIVITY INSTITUTE

International Productivity Monitor

The *International Productivity Monitor* (IPM) is an open-access, peer-reviewed economics journal published by the Centre for the Study of Living Standards in partnership with The Productivity Institute. The *IPM* publishes high-quality articles that build the evidence base and serve as a forum for global knowledge exchange on productivity research. Articles are contributed by leading academics and practitioners, combining rigorous analysis with practical insight, and are intended for a broad audience, including researchers, policy analysts, policymakers, the media, and the general public.

Centre for the Study of Living Standards
170 Laurier Avenue West, Suite 604,
Ottawa, Ontario, K1P 5V5,
Canada
www.csls.ca

The Productivity Institute
Alliance Manchester Business School,
Booth Street West, Manchester, M15 6PB,
United Kingdom
www.productivity.ac.uk

ISSN 1492-9767 on-line

Table of Contents

INTERNATIONAL PRODUCTIVITY MONITOR NUMBER 50, Spring 2026

Editors' Overview

The publication of the 50th issue of the *International Productivity Monitor* (IPM) marks an important milestone for the journal. Since its founding 25 years ago, the *IPM* has now published 350 articles by academics and policy practitioners from around the world. The journal's success reflects the leadership and dedication of Andrew Sharpe. His founding role and longstanding stewardship as Managing Editor established the *IPM* as a respected forum for international knowledge exchange on productivity issues. We look forward to building on the strong foundation he established.

To inform the journal's future direction, we conducted a survey of *IPM* contributors and readers. Respondents rated the journal highly, citing as key strengths its focus, empirical rigor, policy relevance, and international perspective. The survey also identified opportunities to further expand the journal's reach and impact, while preserving its core strengths. These opportunities include several new initiatives that are underway, such as enhancing the dissemination of findings through accompanying blogs, webinars, and podcasts, improving data accessibility, and modernizing the journal's design and website.

This 50th issue contains six contributions covering five broad themes: artificial intelligence, intangible capital, manufacturing productivity, health care measurement, and the high cost of living. **Martin Baily**, **David Byrne**, **Aidan Kane** and **Paul Soto** open the issue with a wide-ranging review article that examines the potential for generative artificial intelligence (GenAI) to provide a sustained productivity boost. They argue that GenAI exhibits characteristics of both a general-purpose technology and an invention in the method of invention. Based on the available evidence, they conclude that GenAI could generate lasting future productivity gains through widespread diffusion, complementary innovations, and more efficient research and development. However, the authors also warn that productivity gains from AI may be delayed or limited by slow adoption, organisational hurdles, and risks of over-investment.

The second article, by **Ahmed Bounfour**, **Kazuma Edamura**, **Takayuki Ishikawa**, **Tsutomu Miyagawa**, **Alberto Nonnis** and **Konomi Tonogi**, analyzes the productivity "J-curve" — the idea that large investments in intangible assets associated with digitalization may result in official statistics underestimating total factor productivity (TFP) early in the investment boom. Examining empirical evidence for five advanced economies, they find that J-curve effects are largely unique to the U.S. and have been much smaller in Europe and Japan. Their estimates highlight the role of software investments in the United States and suggest that standard U.S. TFP growth measures may be underestimated by up to 1.6 percentage points annually. The authors conclude by calling for more aggressive investment in digital innovation in Europe and Japan.

The next two articles, originating from a CSLS session at the 2026 American Economic Association Annual Meeting, examine why U.S. manufacturing productivity growth slowed so dramatically, falling from an annual rate of 3.3 per cent during 1987-2010 to -0.3 per cent during 2010-2023.

Robert Gordon and **Kenneth Ryu** argue that the disappearance of productivity growth after 2010 can be traced back a decade earlier, and partly attributed to the surge in U.S. manufacturing imports that began around 2000 — roughly the time when China joined the World Trade Organization. They show that increased U.S. manufacturing imports came not only from China, but also from other lower-wage economies such as Mexico. They find these imports ultimately weakened domestic manufacturing production, capacity utilization, profitability, investment and innovation activities. Beyond import competition, the authors also point to diminishing returns to innovation, regulatory distortions, and a shortage of skilled labour as contributing factors.

Danial Lashkari and **Jeremy Pearce** further examine the sources of the slowdown by decomposing U.S. manufacturing productivity growth into contributions from leader and follower firms within frontier and laggard industries. They demonstrate that the slowdown was broad-based. It occurred across productivity measures, using multiple industry groupings, and for both leaders and follower firms. The broad-based nature of the slowdown raises the question of whether the translation of research and development (R&D) expenditures into productivity gains has weakened, as suggested by Gordon and Ryu. The authors estimate an R&D production function at the industry and firm levels and find that the elasticity of productivity with respect to R&D is consistently larger in the earlier period than in the full sample period, even as R&D expenditure rose across firms and industries. These results suggest that the slowdown reflects declining research productivity rather than reduced innovation effort and are consistent with the "ideas getting harder to find" hypothesis.

The fifth article, by **Calvin Ackley**, **Abe Dunn**, **Eli Liebman** and **John Romley**, seeks to better understand health care productivity in the United States. Productivity is a critical issue for a sector that accounts for 17 per cent of U.S. GDP. However, official statistics likely understate productivity growth by failing to capture improvements in medical technology and treatment quality. The U.S. Bureau of Economic Analysis has developed a Health Care Satellite Account (HCSA) to try to address this gap by measuring spending by medical condition, thus moving from an input to an output-based measurement in health care. The authors present a framework that combines the HCSA with population health data to adjust output prices for quality improvements. In this framework, "output" is defined as marginal health gains rather than traditional service counts or hours worked. Their results suggest substantial quality-adjusted productivity growth that is largely absent in official statistics. They speculate that health care productivity gains might be even larger in other high-income countries, where life expectancy has increased more than in the United States while health care spending has grown more slowly.

The 50th issue concludes with a commentary by **Claude Lavoie** that studies persistent concerns about the high cost of living among Canadians. The issue is

pervasive, despite macroeconomic indicators showing that household income growth has generally outpaced inflation in recent years. Drawing on polling data, economic statistics, and policy literature, the author identifies housing affordability, slowing real income growth, and social media–driven financial perceptions as key drivers. The commentary argues that policy responses should prioritize housing supply, productivity growth, and stronger worker bargaining power to help ensure that economic gains translate into improved household welfare.

The Potential for Sustained Productivity Impetus from GenAI

Martin Neil Baily, David M. Byrne, Aidan T. Kane and Paul E. Soto*

IPM Review Article

Abstract

With the advent of generative artificial intelligence (GenAI), the scope of AI has increased dramatically, but its effect on labour productivity remains uncertain. Some innovations raise labour productivity growth as adoption spreads but the effect fades when the market is saturated. In contrast, two types of innovation — general purpose technologies (GPTs) and inventions in the method of invention (IMIs) — have long-lasting effects on productivity growth. GPTs (1) are widely adopted, (2) spur knock-on innovations, and (3) show continual improvement, refreshing the innovation cycle. IMIs increase the efficiency of the research and development process via improvements to observation, analysis, communication, and organization. We conclude there is suggestive evidence that GenAI is both a GPT and an IMI, a sign that its adoption will lead to higher labour productivity growth in the future.

*Martin Neil Baily is Senior Fellow Emeritus and Aidan T Kane is a Research Analyst at The Brookings Institution. David M. Byrne and Paul E. Soto are both Principal Economists at the Federal Reserve Board of Governors. David Byrne is the corresponding author (email: david.m.byrne@frb.gov). Authors are listed in alphabetical order, not in order of relative contribution. The views expressed here are not represented to be the views of the staff or trustees of The Brookings Institution nor of the Federal Reserve. The authors are grateful to Michael Chui, Leland Crane, Avi Goldfarb, Bob Gordon, Shane Greenstein, Anton Korinek, James Manyika, Sid Srinivasan, Scott Stern, and Bill Whyman for helpful conversations.

1. Introduction

In late 2022, OpenAI grabbed the world's attention with ChatGPT, a generative artificial intelligence (GenAI) program that uses a computer model of human discourse to respond to natural-language questions (Figures 1-1 and 1-2). The scope of AI expanded dramatically with the advent of GenAI, including to tasks previously seen as quintessentially human, such as competition-level mathematics (Figure 2). Indeed, more and more challenging benchmark tests have been needed to assess the technology's progress, as GenAI has matched human performance on one task after another. In another encouraging sign, field test evidence of productivity improvements from GenAI in practical applications has emerged, notably for writing, computer programming, and responding to call center inquiries. We supplement the quantitative evidence with a qualitative assessment of the properties of GenAI, with a view to helping to predict its impact.[1]

We focus on two particularly impactful types of innovation. Unlike discrete innovations that temporarily raise productivity growth as adoption spreads, "general purpose technologies" (GPTs) and "inventions in the method of invention" (IMIs) have longer-lasting growth effects. GPTs are widely adopted, spur knock-on innovations — new products, process improvements, and business reorganization — and refresh this adoption cycle through ongoing improvement in the core technology (Lipsey *et al.*, 2005). IMIs yield a sustained increase in productivity growth by lowering the cost of research and development (Whitehead, 1925). We conclude there is suggestive evidence that GenAI is both a GPT and an IMI, a sign that its adoption will lead to higher labour productivity growth in the future.[2]

2. GenAI as a General Purpose Technology

The future productivity impact of GenAI will depend on whether (1) it is widely adopted, (2) it spurs related innovations, and (3) the technology continues to improve. These are the three distinguishing features of GPTs.

2.1 Diffusion

Analysis of job descriptions suggests that AI can be used for a broad range of workplace tasks, indicating high potential for diffusion and rising use of AI (Eloundou *et al.*, 2024).[3] We review evidence that this is taking place. Although few of these indicators are restricted to GenAI, the timing

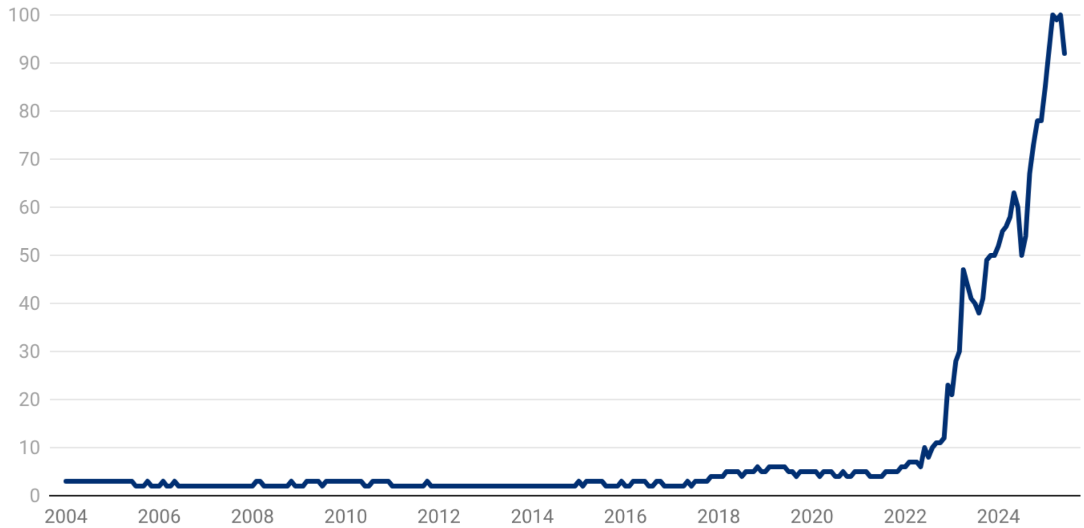
1 For a more extensive discussion of the issues in this article, see Baily and Kane (2025a,b); Kane and Baily (2025a,b).

2 There is a substantial literature on the question of whether machine learning (ML), which preceded GenAI, is a GPT or an IMI, but little such work on GenAI. Eloundou *et al.* (2024), an exception, consider the prospects for GenAI to be a GPT based on the prevalence of tasks that appear likely to benefit from GenAI. On ML as a GPT, see Cockburn, Henderson, and Stern, 2019; Trajtenberg, 2018; Bresnahan, 2019; Goldfarb, Taska, and Teodoridis, 2023; Bresnahan, 2024. Cockburn, Henderson, and Stern (2019) consider if ML is an IMI.

3 Yin, Vu, and Persico (2026) caution that measures of exposure to AI are highly sensitive to the assumptions used.

Figure 1-1: Web Searches for Artificial Intelligence

Indexed to 100 in month of greatest search

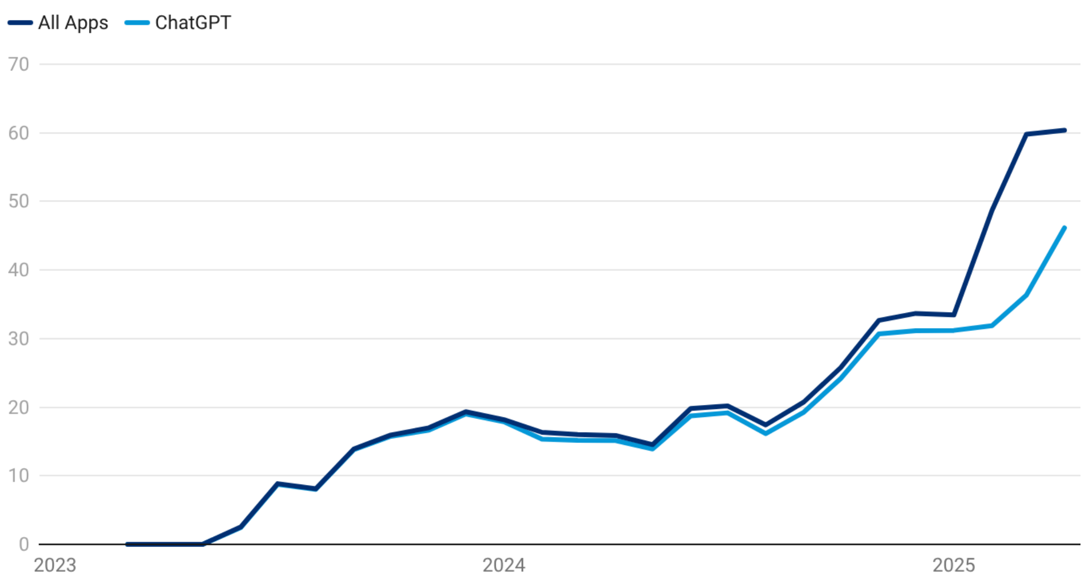


Note: Web searches include related terms in Google's "AI" topic.

Source: Google Trends. • Created with Datawrapper

Figure 1-2: Generative AI Mobile App Downloads

Millions of downloads

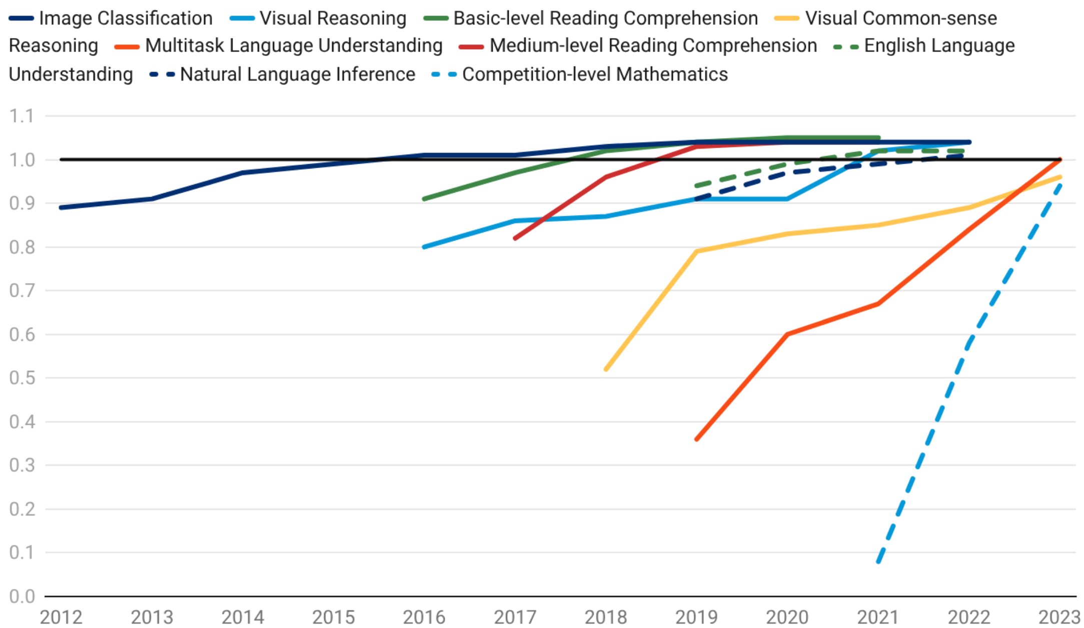


Note: Apps include ChatGPT, Claude, DeepSeek, and Perplexity. Includes Android and iOS versions. Android download information not available for China. Does not account for access via application program interface.

Source: Appfigures. • Created with Datawrapper

Figure 2: Artificial Intelligence Benchmark Performance

Performance relative to human baseline on various tasks, ratio



Note: The human baseline concept used varies by task. For more challenging tasks, the baseline reflects expert-level performance.

Source: Reproduced with permission from the 2024 AI Index Report, Stanford Institute for Human-centered Artificial Intelligence. • Created with Datawrapper

of their increases is consistent with GenAI driving adoption upward. The level of AI use varies across sources and depends on the meaning one uses for adoption.

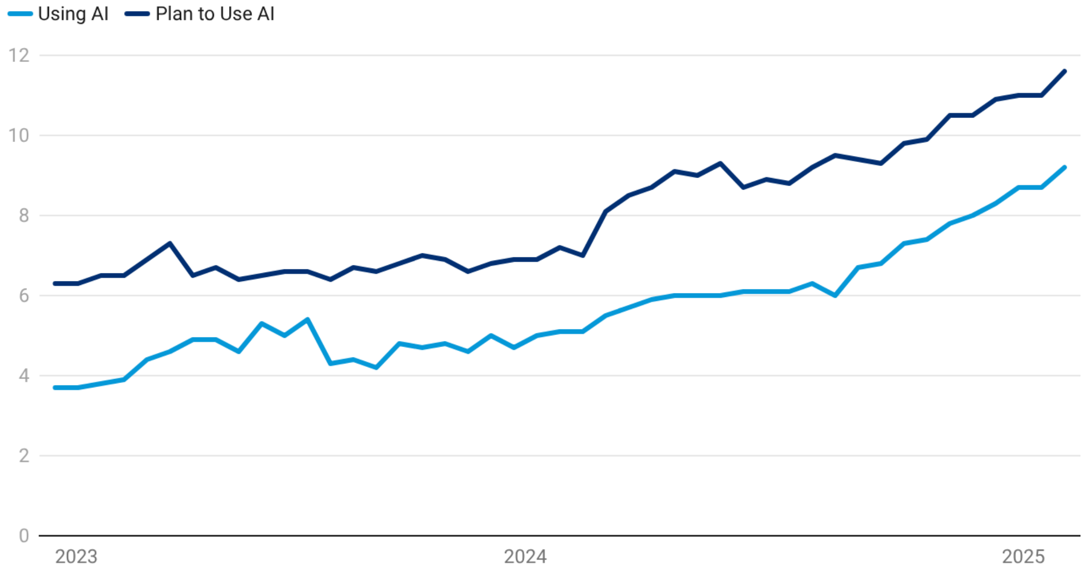
Large-scale firm surveys from the U.S. Census Bureau and McKinsey provide rather different pictures of adoption at first glance (Figures 3-1 and 3-2). The Census Bureau's Business Trends and Outlook Survey (BTOS) found that roughly 17 per cent of firms used AI (GenAI and other types of AI) as of March 2026.[4] In contrast, McKinsey reported that 88 per cent of firms did so in November 2025. Differences in survey coverage likely explain much of the gap. The BTOS is a representative sample of 200,000 U.S. firms, only a handful of which are large corporations. In contrast, the McKinsey survey is a convenience sample with heavy representation from large corporations (McKinsey, 2024). That is, large firms appear to use AI far more than small firms.

Surveys of workers suggest significant AI adoption with relatively few adopters reporting frequent use. Gallup, Inc. reports 50 per cent of U.S. workers had used AI at work as of the fourth quarter of 2025, though only 13 per cent of respondents used it daily. Pew Research Center reports that 21 per cent of U.S. workers used AI for at least some of their work in

4 See Bonney *et al.* (2024) for details about the survey and Bonney *et al.* (2026) for a more comprehensive exploration of its implications.

Figure 3-1: Census Bureau Business Survey (U.S.)

Per cent

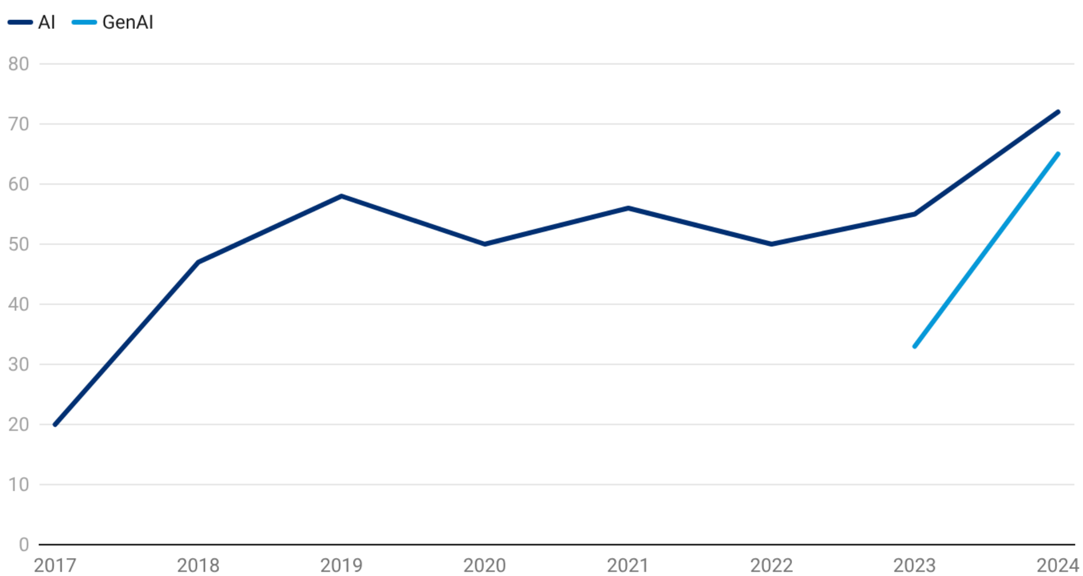


Note: Respondents were asked about AI use in producing goods or services during the past two weeks and anticipated in the next six months.

Source: Census Bureau, Business Trends and Outlook Survey. • Created with Datawrapper

Figure 3-2: McKinsey Survey (Global)

Per cent



Note: Respondents were asked if they "use AI in at least one business function".

Source: McKinsey, "The State of AI in Early 2024." • Created with Datawrapper

September 2025. Bick, Blandin, and Deming (2024) ask specifically about GenAI and find nearly 40 per cent of workers used it in 2024. As new workers enter the labour force, these numbers are expected to rise. The Pew poll finds that 64 per cent of teens have used an AI chatbot and more than half of them have used them to search for information and to get help with schoolwork.

AI use varies significantly across industries. Both the BTOS and the McKinsey survey show that use is particularly high in the information sector, where computer coding is prevalent. While our focus in this article is not the labour market, we note two indicators of the disruption caused by AI in the information sector. Data from Lightcast show that the share of job postings in the information sector mentioning AI and related skills was 31 per cent in May 2026, up from 10 per cent in 2021, and far higher than the average across all sectors of 7 per cent (Figure 4). Crane and Soto (2026) provide evidence that labour demand growth for computer coders slowed noticeably following the release of ChatGPT.[5]

Kane and Baily (2025a,b) provide industry case studies summarizing the evidence of GenAI adoption in four sectors — information, healthcare, finance and electricity generation and distribution. In some examples, such as computer coding, GenAI has been adopted quickly, and in all these sectors companies were experimenting with ways to cut costs using GenAI. There are often barriers to adoption, however, including regulation, skill shortages, and high adoption costs. In other cases, efforts to use GenAI have proven disappointing so far.

2.2 Knock-on Innovation

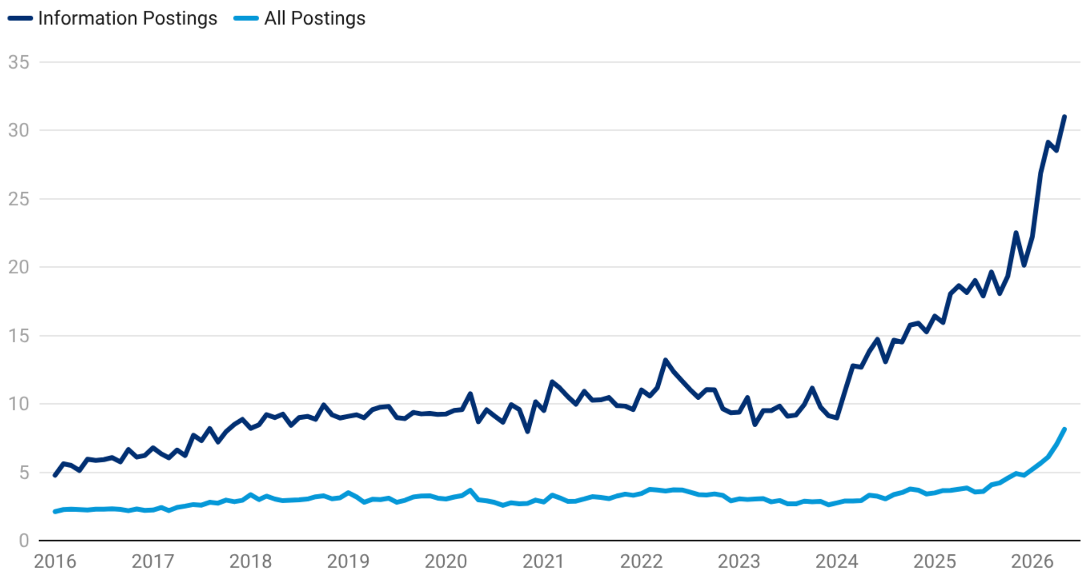
Knock-on innovations related to GenAI include novel and improved software, more efficient product and process design systems, and organizational changes to better exploit GenAI capabilities.

A prominent example of knock-on innovation is ChatGPT itself, a user interface (UI) for OpenAI's GenAI models. UIs provide a channel through which requests and responses can pass between the GenAI model and human users. ChatGPT, released in 2022, is a conversational interface that made GenAI interactions significantly more accessible relative to early approaches that relied on Python programs or websites such as the OpenAI Playground. In 2023, OpenAI introduced custom GPTs, enabling users to create domain-specific large language models (LLMs), such as LegalGPT for legal matters (OpenAI, 2023). In 2024, OpenAI announced integration of their ChatGPT model to Apple's Siri voice assistant and Google launched NotebookLM, which made it easy to upload documents and transform them into interactive discussions (Johnson, 2024). In addition, there are "copilots" that integrate AI into existing user workstreams, notably GitHub Copilot (computer programming) and Microsoft 365 Copilot (office productivity).

5 On the labour market effects of AI, see Acemoglu *et al.* (2020), Brynjolfsson, Mitchell, and Rock (2018), Felten, Raj, and Seamans (2019), Webb (2019), Eloundou *et al.* (2024).

Figure 4: AI-Related Job Postings

Per cent of total



Note: Jobs classified using AI-related terms found in job descriptions, as described in Acemoglu et al. (2020). List of AI-related terms updated by Lightcast.

Source: Lightcast. • Created with Datawrapper

System interfaces are another key locus of knock-on innovation for GenAI. These allow hardware and software systems to access the AI model. For example, Nvidia's Isaac Software Development Kit (SDK) facilitates the integration of AI into robotics. Access to AI through the SDK helps the robot with environmental integration problems, such as simultaneously tracking its location and mapping its environment. Development of multimodal models — which can take in inputs of different kinds (text, images, sensor readings) and output instructions to the robot, such as the rotation and torque for a joint — has pushed robot-AI integration forward (Reed *et al.*, 2022; Brohan *et al.*, 2023). System interfaces also enable the design of agentic AI systems, which collect information during operation, interpret context, make decisions and act autonomously to pursue goals (Park *et al.*, 2023). Innovations in production line operations have followed GenAI advances as well. Serradilla *et al.* (2022) provide an overview of the use of deep learning, including GenAI such as generative adversarial networks, to optimize line configuration, throughput, efficiency, and carbon footprint. Predictive maintenance using synthetic data and scenario simulation is another application for GenAI in industry.

Another area of knock-on innovation has been the creation of AI agents. Agentic AI systems develop strategies to pursue broad goals and recalibrate in response to their environment, in contrast to tool-based AI, which has a stable structure and calibration, and is only equipped to respond to carefully crafted requests. For example, specialized agents can handle health care paperwork including making appointments, following up on treatment protocols and submitting insurance claims. In software development, agentic tools have extended the copilot approach further. Tools such as Anthropic's *Claude Code* allow programmers to delegate command-line coding tasks, such as debugging, refactoring, and test execution, with the agent reading and modifying files autonomously rather than suggesting individual lines of code.

2.3 Core Innovation

Importantly, labour productivity (economic performance, as opposed to benchmark performance) only rises when more can be accomplished *while holding input costs fixed*. Accordingly, we focus below on (a) how innovations in model architecture raise GenAI model capabilities without raising training costs, (b) how hardware innovations lower the cost of computation, and (c) how richer datasets can be brought to bear on training.

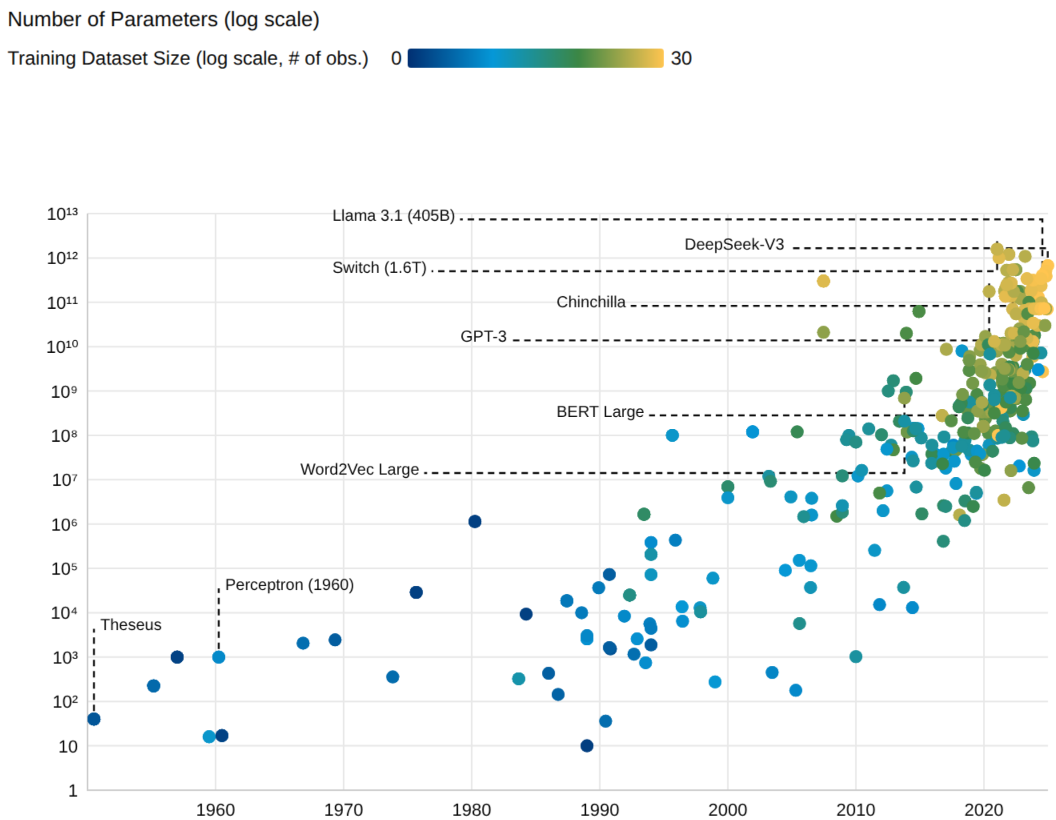
2.3.1 Model Development, Training, and Deployment

Early in the development of modern GenAI, advances in performance were attained primarily by increasing model scale (number of parameters), computational power, and training data (Kaplan *et al.*, 2020). That is, output (benchmark performance) was increased by adding more inputs (Figure 5). For example, GPT models had 117 million parameters in 2018 (GPT-1) and as many as 175 billion in 2020 (GPT-3) (Shree, 2020). More recently, increasing attention has been given to using these inputs more efficiently. These efficiency gains appear in four distinct stages: leveraging insights into algorithm design to make better models, improving model pre-training, fine-tuning of those models, and strategies to reduce the cost of inference (running the deployed models).

The Transformer, which enables GenAI to efficiently pay attention to context when interpreting text, was the seminal algorithmic innovation for modern GenAI (see Box 1). Further progress on this "attention mechanism" has followed. An example is Mamba (Gu and Dao, 2023). In the original transformer, computational burden is proportional to the square of the number of parameters; Mamba achieves sub-quadratic costs using a state-space model. Other innovations have focused on what can be achieved with smaller scale models; models from Microsoft and Mistral AI have shown strong performance relative to their size (Jiang *et al.*, 2023; Abdin *et al.*, 2024). Open-source model development has played a central role in this process.

Pre-training brings algorithms to a dataset to produce a broadly applicable "foundation model." Between 2018 and 2022, a key pre-training tactic in the effort to improve GenAI performance was to increase model size. Remarkably, algorithmic innovation has more than offset the resulting increase in computational complexity: the cost of training a model of a given size was halved approximately every eight months through 2024 (Ho *et al.*, 2024). This focus on scale abated to some degree over time: by 2022, diminishing returns to model size for foundation models had appeared and researchers have increasingly explored performance improvements in fine-tuning and inference (Zeff, 2024).

Figure 5: Number of Parameters and Training Dataset Size



Note: Only models with reported parameter size and training dataset size are included in the dataset. For instance, GPT-4 is excluded as its parameter size is not known.
Source: Epoch (2024) with major processing by Our World in Data (Rahman, Owen, and You 2024).

Fine-tuning refines the foundation model for a specific application. Fine-tuning innovations have included transfer learning (adapting an already fine-tuned model to a related task with domain-specific data), instruction tuning (guiding the model to recognize instructions, not just predict the next word; Taori *et al.*, 2023), and reinforcement learning from human feedback (aligning the model outputs with human preferences; Christiano *et al.*, 2017; Ouyang *et al.*, 2022). Once pre-trained and fine-tuned, the model is used in inference (responding to user requests). The aggregate cost of GenAI inference — in terms of electricity, time, computation, and carbon emissions — has risen with the popularity of GenAI, leading to a focus on techniques to make this step more efficient.[6]

Conversely, some recent models have deliberately extended inference time to improve performance with such techniques

6 Among these innovations are the Mixture of Experts (MoE) approach — only activating a subset of model parameters (Jacobs *et al.*, 1991; Shazeer *et al.*, 2017); pruning — removing extraneous parameters (Cetin *et al.*, 2024); distillation — compressing large models into smaller ones (Hinton, Vinyals, and Dean, 2015); quantization — reducing numerical precision (Wang *et al.*, 2023); and token caching — storing reusable computations (Pope *et al.*, 2023).

Box 1: The Transformer

The transformer architecture, introduced by Vaswani *et al.* (2017), was a game-changer in AI, particularly as the engine behind GenAI models. Its key innovation, the "attention mechanism," steers models to focus selectively on relevant parts of the prompt, enabling more efficient and accurate processing of language. This breakthrough has powered major advancements in natural language understanding, translation, and generation, forming the backbone of today's most advanced GenAI systems.

Transformers process input data through a series of layers (steps), each consisting of an attention mechanism followed by a multilayer perceptron (MLP, defined below), proceeding as follows.

First, a representation of the prompt (input text) suitable for analysis by the model is created. Specifically, the prompt is broken into tokens (smaller pieces which may be phrases, words, or parts of words). The tokens are converted into embeddings (numerical vector representations) which encode the semantic and syntactic meaning of each token. Loosely speaking, for each token, the closest of the other tokens, as measured by the distance between their embeddings, are the ones most important to understanding its meaning.

Second, the attention mechanism processes the matrix of token embeddings using three large matrices called the "query," the "key," and the "value." For each token in the input, the query is compared to the keys of all tokens to compute attention scores, which are used to form a weighted average of values. This step allows each token's representation to incorporate information from other tokens in the prompt based on their contextual relevance.

Third, the data passes through an MLP, a type of neural network. While the attention mechanism focuses on pairwise interactions between tokens, the MLP applies nonlinear functions (in contrast to the linear attention mechanism) in refining the token representations.

This sequence—of computing the attention mechanism followed by the MLP—is repeated multiple times depending on how many layers are in the model (for example, the Llama-3 model has 32 layers), enabling the model to capture increasingly abstract features of the input text.

The performance gains from scaling of this system through increasing the size of these matrices—along with larger training datasets and improvements in hardware and processing algorithms—underpins the rising ability to handle complex language tasks.

as chain-of-thought reasoning (Wei *et al.*, 2022). A salient example of the cumulative effect of these innovations is DeepSeek R1, released in January 2025, which blended several of these approaches, including MoE, chain-of-thought reasoning, and reinforcement learning and distillation, to achieve frontier-level performance at a fraction of the cost of comparable models (DeepSeek-AI, 2024).

2.3.2 Hardware

In addition to the software and operational choices described above, GenAI model training and inference costs rely critically on the state of electronic hardware. Because the workloads generated by GenAI training are best handled with massive parallel computation, graphics processing units (GPUs), which are designed for parallel processing, play a central role.[7] Successive GPUs released by NVIDIA have delivered leaps in AI performance through improvements in processing core (CUDA) design, the addition of on-chip tensor cores, which accelerate matrix calculations, and massive amounts of high-speed integrated memory. In addition to these circuit design innovations, hardware cost improvement depends on advances in basic and applied science that permit greater miniaturization and power-efficiency for electronics.[8] While the engineering performance of leading-edge GPUs has rocketed upwards in recent years, prices have increased dramatically as well, but by less than performance: In 2007, a $349 GPU provided 0.3 teraflops (TFLOPS) of compute and in 2024, a $299 GPU delivered 15.1 TFLOPS, implying an average annual rate of price decline of 24 per cent that persisted for 17 years (Figure 6). This example is representative of a broader trend.

More recently, NVIDIA's Blackwell GPU architecture, introduced in 2024, roughly quadrupled the inference throughput of LLMs (tokens per second) at comparable power consumption, relative to prior models like the Hopper. Such hardware gains amplify the algorithmic efficiency improvements described above for training as well. Given that agentic AI systems involve long, multi-turn interactions in which input text is reused across steps, these types of gains are particularly consequential for the rise of agents discussed in Section 2.2, where inference costs rather than training costs are binding constraints on adoption.

2.3.3 Datasets

GenAI models "learn" by adjusting parameters to best represent the content of large amounts of text, allowing them to choose the word that should appear next in response to a user's prompt. Figure 5 illustrates the increase over time in the size of the datasets used in training and the associated improvement in (benchmark) performance. A crucial nuance to this aspect of model improvement is that access to more information, not more text, is needed to continue to improve GenAI models. Diminishing marginal returns set in as developers move from information-rich content, such as Wikipedia and scientific articles to noisier text like social media posts.

One approach to mitigating the content constraint is transfer learning, where a model pre-trained with public data is improved by further training using proprietary data. Another approach is data augmentation, such as incorporating small, localized modifications of the training data. For example, the performance of an image recognition model may be improved by

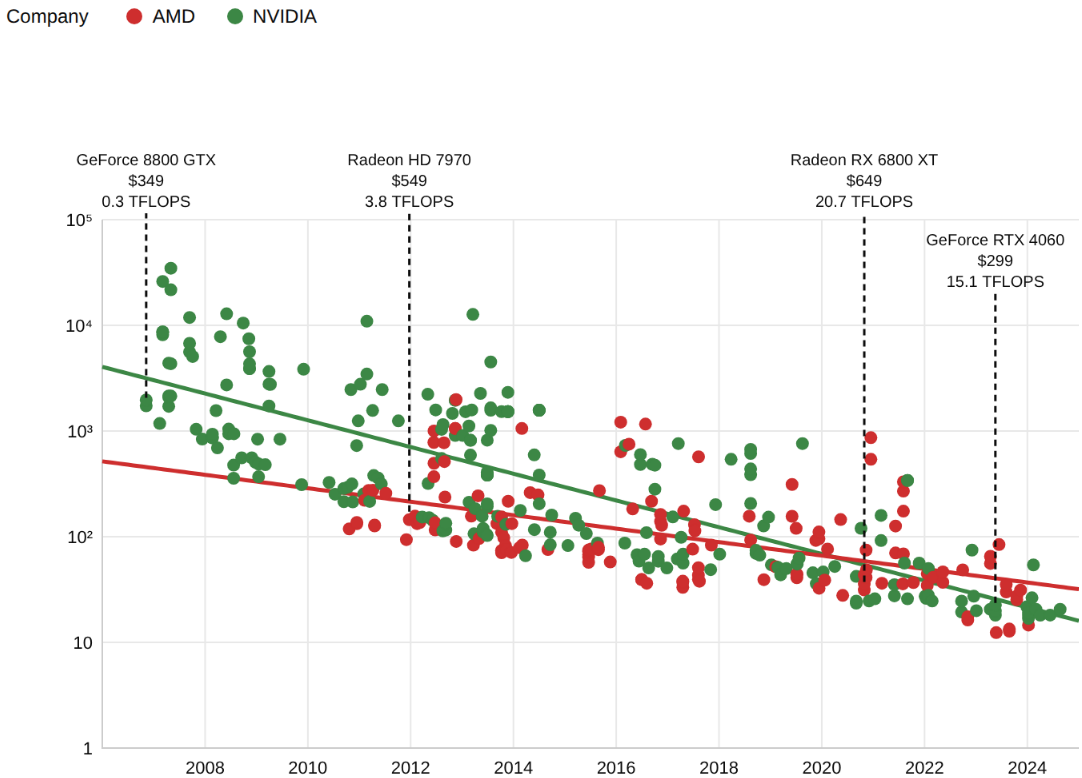
7 Specialized chips, such as tensor processing units (TPUs) customized for matrix multiplication, have increased computational efficiency as well.

8 The frequent release of new generations of semiconductors belies the difficult challenges faced in achieving each one. The Institute of Electrical and Electronics Engineers regularly publishes "roadmaps" detailing the problems that must be solved for continued improvement in electronics performance (see International Roadmap for Devices and Systems (IRDS™) 2024 Edition https://irds.ieee.org/editions/irds2024/.)

9 This approach was taken by the developers of AlexNet, a model which revolutionized the field (Krizhevsky, Sutskever, and Hinton, 2012).

Figure 6: Price Trends for GPUs, Price per TFLOP

Dollars per Trillion Floating-Point Operations per Second (log scale)



Note: The green line represents the line of best fit for NVIDIA, and the red line represents the line of best fit for AMD.

Source: TechPowerUp.

supplementing the training set of labeled images with their mirror images.[9] "Synthetic data", where generative models create new data to emulate patterns and characteristics of real data has been explored as well (Liu *et al.*, 2024).[10] Last, datasets can be augmented by harvesting information collected with sensors, particularly in physical environments such as industrial robots and autonomous vehicles (Feng *et al.*, 2019).

3. GenAI as Invention in the Method of Invention

Like other sectors, efficiency in the research sector can be increased using appropriate capital, such as inventions in the method of invention (IMI). We consider below whether GenAI is such an invention and whether it can contribute to research productivity beyond what is contributed by predecessor AI technologies. We then review broad indicators of AI's role in

10 Some observers have raised concerns that training with synthetic data (and AI-generated text increasingly present on the internet) will yield low-quality or even nonsensical results, a phenomenon known as "model collapse" (Alemohammad *et al.*, 2023; Shumailov *et al.*, 2023). Others have argued that model collapse only occurs when the original training text is replaced by model-generated text (Gerstgrasser *et al.*, 2024).

research: patent filings, the share of AI use by workers in research roles, and new evidence on AI references in company conference calls.

Eloundou *et al.* (2024) note that "scientists and researchers" and "technologists" are the job groups most highly exposed to LLMs, suggesting substantial potential for research and development (R&D) productivity enhancement from GenAI. Prior to GenAI, AI had already diffused widely across scientific disciplines and improved research efficiency (Carobene *et al.*, 2024). Cockburn, Henderson, and Stern (2019; 2023) note that pre-generative AI assists with the "labour-intensive search with high marginal cost of search" involved in many types of R&D. Agrawal *et al.* (2018), emphasize that AI improves prediction, which plays a central role in research; for example, machine learning has also been used extensively for predicting the properties of novel metal alloys, economizing on physical experimentation and computer simulations (Hart *et al.*, 2021). Our focus is on the question of whether GenAI enables additional efficiencies in R&D beyond these and other improvements provided by machine learning. We group IMIs into enhancements to observation, analysis, communication, and organization.

3.1 Observational

Observational tools, such as microscopes, telescopes, and cameras are central to scientific advance (Mokyr, 2004). These tools are invariably limited in that they produce imperfect data due to defects in their components and variation in the environment. GenAI provides a tool to enhance imperfect portions of data. For example, generative techniques for image enhancement perform better than techniques, such as splines, relying solely on smoothness assumptions (i.e., that nature does not make leaps) (Liu *et al.*, 2018; Lugmayr *et al.*, 2022). GenAI can approximate the manifold of the data generating process, implicitly accounting for the actual physics, say, of a remote galaxy seen through an imperfect lens. More generally, imputation of missing observations in datasets of all kinds is possible in a fashion consistent with the apparent properties of the underlying phenomena.

3.2 Analytical

GenAI can be used as an analytical tool as well. Caliskan, Bryson, and Narayanan (2017), for example, find that "text corpora contain recoverable and accurate imprints of our historic biases." This new visibility may promote analysis of previously intractable social science questions. There has been an explosion of sentiment analysis and other forms of natural language processing in recent years fueled by this capability of GenAI.[11] While the identification of underlying sentiment (encoding) is strictly speaking, a function of the LLM, conveying the discovered sentiment to the user is necessarily a generative process. Korinek (2023) documents a variety of potential roles for GenAI in the economic

11 Sentiment analysis is possible with earlier forms of AI but the capabilities of GenAI models are vastly greater (Gentzkow, Kelly, and Taddy, 2019; Dell, 2025).

research process; that GenAI may play a similar role in many other fields is a reasonable conjecture.

3.3 Organizational

The organization of science may benefit from the use of GenAI. Institutional organization plays a central role in the effectiveness of R&D (Mowery and Rosenberg, 1999), as do informal associations into professional networks (Wang and Barabasi, 2021) and geographic clusters (Porter and Stern, 2001). Consequently, the method of invention for any given research program properly includes the institutions involved. Emerging applications of AI "digital twins" offer the prospect of R&D with a reduced institutional footprint in many areas of study. Among these are drug discovery (Bordukova *et al.*, 2024), industrial research (Tao, Zhang, and Zhang, 2024), and materials science (Kalidindi *et al.*, 2022). GenAI tools can help with applied science as well, such as designing products that meet technical and aesthetic specifications. Moreover, the design process itself can be transformed to create a range of options, not just one-off designs, together with detailed manufacturing specifications (Saadi and Yang, 2023).

3.4 Communication

Perhaps most obviously, GenAI is a communication tool. Although empirical and analytical stages of research projects focus on measurement and calculation, many aspects of the research process involve manipulating language. GenAI may be employed in the writing tasks involved in the conceptual, planning, and dissemination stages of research projects, such as drafting literature reviews, grant applications, and seminar slides. Whether GenAI improves the efficiency of such tasks on net, once the effort needed for review and editing of the documents drafted by GenAI is accounted for, is an open question. If so, GenAI may play a similar role to the printing press and word processing as a catalyst to the invention process.

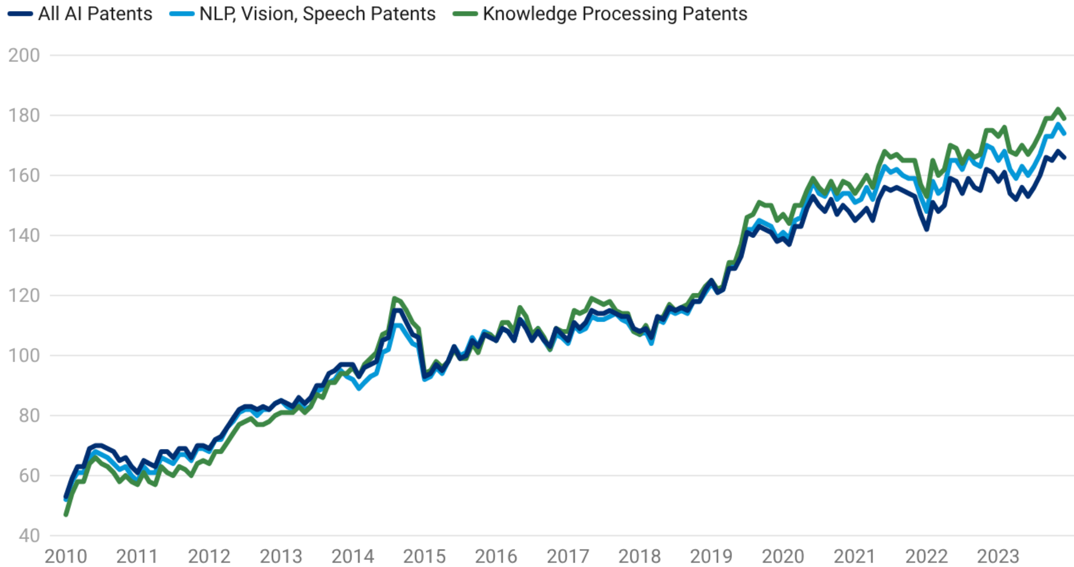
3.5 Indicators of GenAI Research and of GenAI Use in Research

Substantial suggestive evidence has emerged that GenAI has enhanced research performance. AI-related patents issued by the United States Patent and Trademark Office (USPTO) increased following the advent of GenAI, suggesting a related surge in GenAI research (Figure 7) (Pairolero, 2025). The USPTO index of AI-related patents began climbing in 2018, shortly after the publication of the seminal paper by Vaswani *et al.* (2017) quickly reaching a level 50 per cent higher, which it has sustained since 2019. We also observe that increases in patent activity for AI modalities particularly related to GenAI — natural language processing, vision, speech, and knowledge processing — have risen even further. This suggests that the recent surge in patenting activity is not merely a reflection of advancements in machine learning.

Handa *et al.* (2025) provide a rich set of information on GenAI use in their Anthropic Economic Index (AEI), assigning millions of conversations from Claude (Anthropic's premier GenAI system) to roughly 3,500 of the tasks defined by the U.S. Department of Labor's O*NET

Figure 7: Artificial Intelligence Mentions in Scientific Patents

Indexed to 100 in 2015, three month moving average



Source: Artificial Intelligence Patent Dataset (2023), U.S. Patent Office. • Created with Datawrapper

Dataset. Table 1 shows the estimated share of prompts accounted for by occupational groups, their employment share, and the ratio of the two. (If prompts were equally distributed across all workers, these ratios would each be equal to 1.) "Computer and mathematical occupations", which includes the computer programmers for whom GenAI use is especially intense, have the highest ratio of prevalence of GenAI use to occupational prevalence and use intensity is nearly as high among scientists. Other occupational groups with high relative prevalence of GenAI use include "arts, design, sports, entertainment and media"; "architecture and engineering"; and "educational instruction and library". The remaining 87.6 per cent of employment is accounted for by occupations which AEI found had a share of Claude prompts roughly equal to or lower than their share of employment, highlighting the concentrated nature of GenAI adoption in the economy at present.

Figure 8 illustrates significant automation and augmentation of tasks among our groupings of research occupations: programmers exhibit the highest automation rate, with over half of the requests handled by GenAI being automation tasks. Social science researchers show slightly lower automation rates, with economists showing over 23 per cent of their prompts being automation focused. Notably, for hard science researchers (e.g., physicists, biochemists), the share of their GenAI use for automation is nearly 15 per cent higher than their natural science counterparts. This difference likely reflects AI's strength in data-intensive and simulation-based research such as those found in hard sciences like physics and materials science.

Firm communication also reveals GenAI's integration into the invention process. Figure 9 plots the number of firms referencing AI in the context of research,

Table 1: Occupations with High Generative AI Use

Job Type	Prompt Share	Employment Share	Ratio
Computer and mathematical	37.2	3.4	10.9
Arts, design, sports, entertainment, and media	10.3	1.4	7.4
Life, physical, and social science	6.4	0.9	7.1
Architecture and engineering	4.5	1.7	2.6
Educational instruction and library	9.3	5.8	1.6
Memo: Other occupations	31.8	87.6	0.4

Note: Percent share of prompts submitted to Claude AI and linked to tasks by Anthropic. Task weights are apportioned equally to all occupations which include that task in O*NET.

Source: Anthropic Economic Index. • Created with Datawrapper

as indicated by the firms mentioning an AI-specific term ("machine learning," "deep learning," "artificial intelligence," "GenAI", or "generative AI") within a research-related context (within 10 words of "inventi-", "research-", or "discover"). A sudden rise appears in 2023, with approximately 60 public companies per quarter mentioning such usage. This increased integration of AI with R&D illustrates the role it plays in corporate innovation.

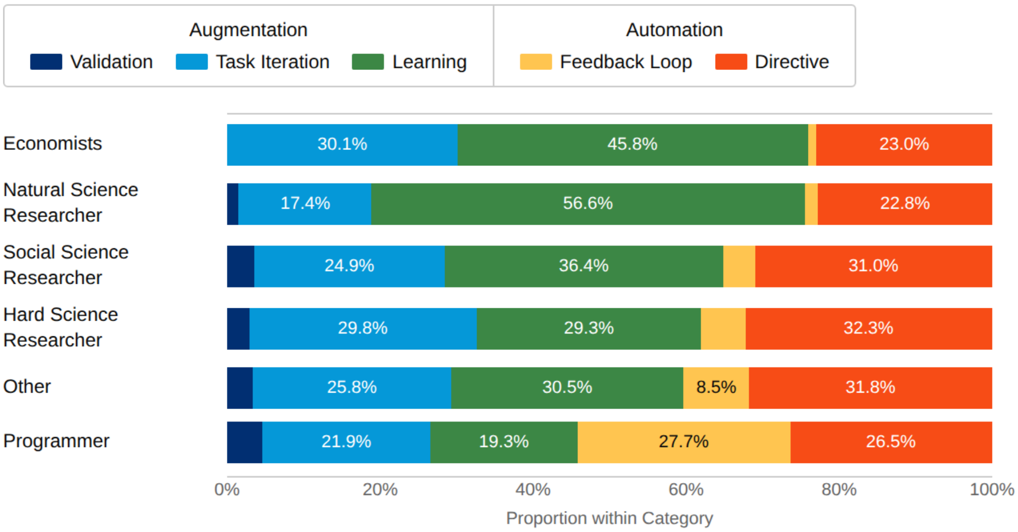
4. Tailwinds and Headwinds for Productivity Growth from GenAI

The qualities of GenAI and the limited evidence on its application suggest that two substantial tailwinds support a forecast of a noteworthy increase in productivity from the technology. GenAI has features typical of both a GPT — headed toward being widely used, stimulating related innovation, and displaying ongoing improvement in (economic) performance — and an IMI — raising the efficiency of R&D through improvements to observation, analysis, communication, or organization. Because both GPTs and IMIs promote productivity growth for extended periods, it is reasonable to expect GenAI will have a noteworthy impact on productivity. That being said, we note several headwinds that should be taken into account.

First, whether the organizational change needed for GenAI to be a true GPT will take place is an open question. AI systems that preceded GenAI demonstrated the need for cross-functional teams with access to data that spans the enterprise, breaking down barriers between business units, optimizing supply chains, and reallocating employees to de-emphasize repetitive writing tasks (Iansiti and Lakhani, 2020). Bresnahan (2024) observed that adoption was concentrated in places where

Figure 8: GenAI Automation vs. Augmentation in Researcher Roles

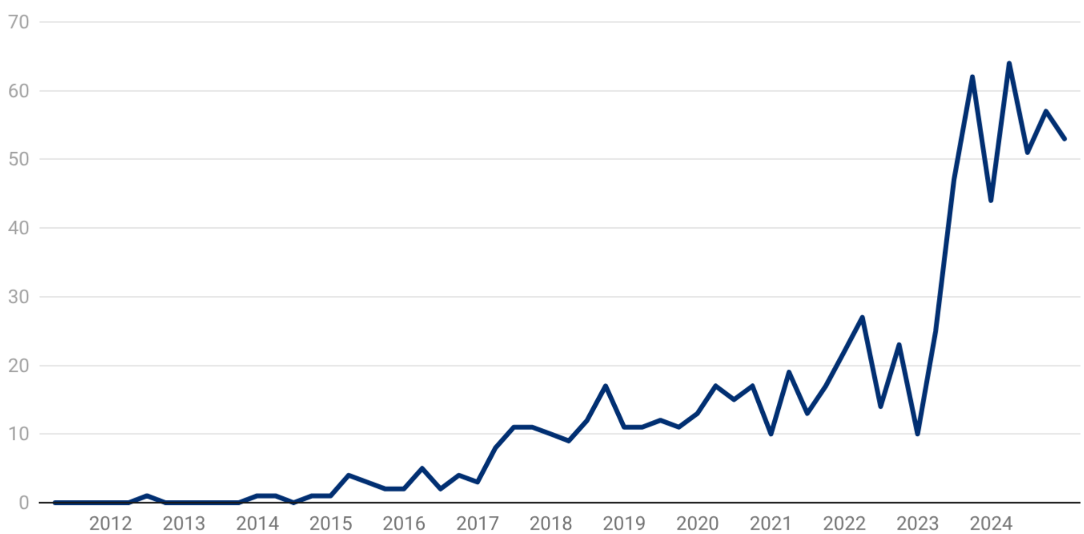
Proportion of AI interactions within each job category, by interaction type



Source: Authors' calculations using the Anthropic Economic Index.

Figure 9: Mentions of Artificial Intelligence Use for Research in Conference Calls

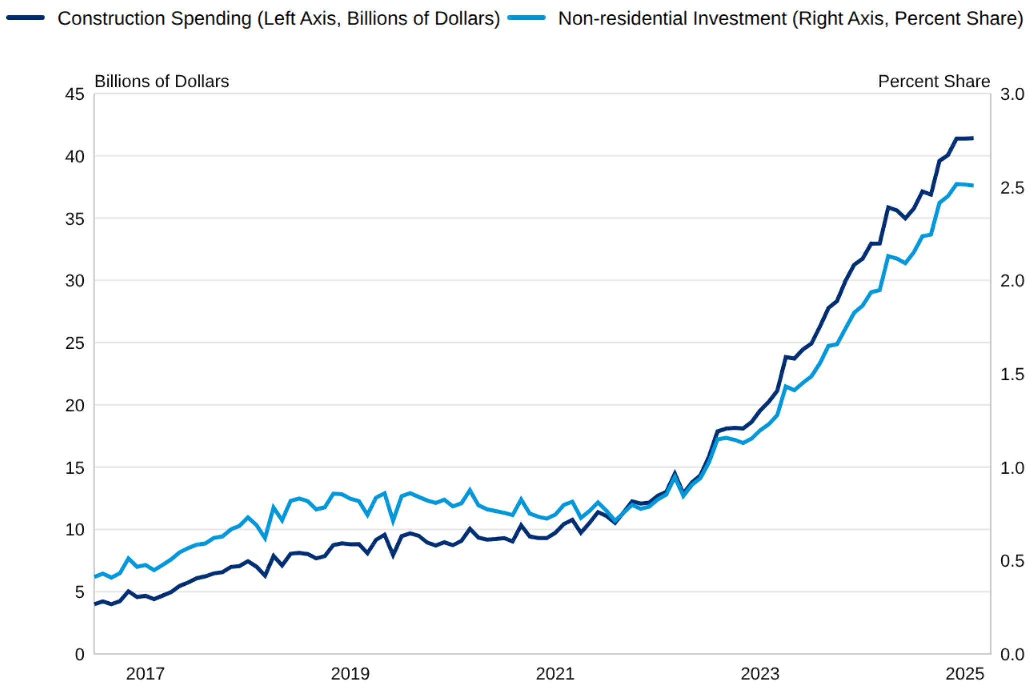
Number of firms mentioning AI and research



Note: Authors' estimates. Data through Q4 2024.

Source: S&P Capital database merged with Compustat. • Created with Datawrapper

Figure 10: Data Center Construction



Note: Excludes IT equipment. Nominal value of construction put in place.

Source: U.S. Census Bureau.

complementary innovation was less necessary, such as in firms that were highly digitized from their founding. These digital natives will surely lead the charge for GenAI as well. For other firms, the pace and success of reorganization innovation will be a key determinant of the scale and timing of productivity effects from GenAI.

Relatedly, the reliance of GPTs on complementary investment tends to damp the effect on labour productivity growth. For example, the effect on the productivity *level* of solid-state computing was large, but it played out over decades. Massive advances in computational technology, including the invention of the solid-state transistor and the fundamentals of system design had accumulated by the end of the 1940s and a steady decline in computing costs had begun (Nordhaus, 2007). The surge in productivity attributed to information technology arrived some fifty years later.

Third, investment to develop and deploy new technologies is fraught with risk. If GenAI is a widely adopted "killer app" that defines a new era of IT, the computing capacity needed to deliver GenAI to millions of simultaneous users will be massive. Anticipation of this outcome helps explain the recent wave of irreversible investment in data centers and power generation (Figure 10). Historically, when such forecasts have proven wrong, the negative consequences of the resulting capital overhang have been substantial.

Fourth, the scope of application of GenAI as an IMI remains to be seen as well. For example, whether GenAI can uncover fundamental features of phenomena is a matter of some debate. Li *et al.* (2022) present evidence that GenAI does develop such knowledge in an "emergent world model." Others argue that GenAI is employing a "bag of heuristics."[12] This question is a crucial one in determining the capabilities of GenAI to contribute to science. Without a model of underlying structure, one cannot articulate fundamental laws. This limitation may be inherent to how GenAI is trained: humans learn scientific fundamentals from textbooks, but such laws may not form the rhetorical backbone of the verbal exchanges that dominate training corpora.

Fifth, we expect that GenAI will boost productivity growth *relative to the counterfactual* economy without it, but the growth effect of machine learning (and other IT innovations) may be waning. The impact of GenAI will have to match the impact of machine learning for the economy simply to match the recent history of productivity growth. In other words, the digital revolution may be baked into the productivity trend and GenAI is just its latest form.

5. Conclusion

The release of ChatGPT in late 2022 was a stark inflection point in public interest in GenAI and predictions of a first-order impact on productivity in the future soon followed, but its economic effects remain uncertain. To complement the limited empirical evidence, we ask what the characteristics of GenAI suggest its future impact on productivity may be. We conclude there is strong evidence that GenAI has the potential to be both a GPT and an IMI. We therefore expect a noteworthy increase in labour productivity from GenAI, though the headwinds we cite suggest the range of plausible outcomes is wide with respect to both the magnitude and timing of the increase.

References

Abdin, M., J. Aneja, H. Awadalla, et al. (2024) "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone," arXiv Preprint arXiv:2404.14219. Available at: https://arxiv.org/abs/2404.14219.

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2020) "AI and Jobs: Evidence from Online Vacancies," NBER Working Paper 28257, National Bureau of Economic Research.

Acemoglu, D. and P. Restrepo (2020) "The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand," *Cambridge Journal of Regions, Economy and Society*, Vol. 13, No. 1, pp. 25–35.

Agrawal, A., J. Gans, and A. Goldfarb (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, Boston.

Alemohammad, S., J. Casco-Rodriguez, L. Luzi, A.I. Humayun, H. Babaei, D. LeJeune, A. Siahkoohi, and R. Baraniuk (2023) "Self-consuming Generative Models Go Mad," in *The Twelfth International Conference on Learning Representations*.

Baily, M.N., D.M. Byrne, A.T. Kane, and P.E. Soto (2025) "Generative AI at the Crossroads: Light Bulb, Dynamo, or Microscope?" Working Paper, Center on Regulation and Markets, Brookings Institution.

Baily, M.N. and A. Kane (2025a) "AI in the Finance Sector: Transforming Productivity and Risk Management," Brookings Institution.

12 A useful entry point to this ongoing debate is "LLMs and World Models," by Melanie Mitchell, February 13, 2025, found at the AI: A Guide for Thinking Humans Substack blog.

Baily, M.N. and A. Kane (2025b) "AI in the Healthcare Sector: Enhancing Care, Efficiency, and Innovation," Brookings Institution. Available at: https://www.brookings.edu/articles/harnessing-ai-for-economic-growth/.

Bick, A., A. Blandin, and D.J. Deming (2024) "The Rapid Adoption of Generative AI," NBER Working Paper 32966, National Bureau of Economic Research.

Bonney, K., C. Breaux, C. Buffington, E. Dinlersoz, L. Foster, N. Goldschlag, J. Haltiwanger, Z. Kroff, and K. 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.

Bonney, K., C. Breaux, E. Dinlersoz, L. Foster, J. Haltiwanger, and A. Pande (2026) "The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks," NBER Working Paper 35141, National Bureau of Economic Research.

Bordukova, M., N. Makarov, R. Rodriguez-Esteban, F. Schmich, and M.P. Menden (2024) "Generative Artificial Intelligence Empowers Digital Twins in Drug Discovery and Clinical Trials," *Expert Opinion on Drug Discovery*, Vol. 19, No. 1, pp. 33–42.

Bresnahan, T. (2019) "Artificial Intelligence Technologies and Aggregate Growth Prospects," in J.W. Diamond and G.R. Zodrow (eds.), *Prospects for Economic Growth in the United States*, Cambridge University Press, Cambridge, England, pp. 132–172.

Bresnahan, T. (2024) "What Innovation Paths for AI to become a GPT?" *Journal of Economics & Management Strategy*, Vol. 33, No. 2, pp. 305–316.

Brohan, A., N. Brown, J. Carbajal, et al. (2023) "Rt-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," arXiv Preprint arXiv:2307.15818.

Brynjolfsson, E., T. Mitchell, and D. Rock (2018) "What can Machines Learn and What Does it Mean for Occupations and the Economy?" *AEA Papers and Proceedings*, Vol. 108, pp. 43–47.

Caliskan, A., J.J. Bryson, and A. Narayanan (2017) "Semantics Derived Automatically from Language Corpora contain Human-like Biases," *Science*, Vol. 356, No. 6334, pp. 183–186.

Carobene, A., A. Padoan, F. Cabitza, G. Banfi, and M. Plebani (2024) "Rising Adoption of Artificial Intelligence in Scientific Publishing: Evaluating the Role, Risks, and Ethical Implications in Paper Drafting and Review Process," *Clinical Chemistry and Laboratory Medicine (CCLM)*, Vol. 62, No. 5, pp. 835–843.

Cetin, E., Q. Sun, T. Zhao, and Y. Tang (2024) "An Evolved Universal Transformer Memory," arXiv Preprint arXiv:2410.13166.

Christiano, P.F., J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei (2017) "Deep Reinforcement Learning from Human Preferences," *Advances in Neural Information Processing Systems*, Vol. 30.

Cockburn, I.M., R. Henderson, and S. Stern (2019) "The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis," in A. Agrawal, J. Gans, and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago, pp. 115–146.

Crane, L.D. and P.E. Soto (2026) "AI and Coder Employment: Compiling the Evidence," Federal Reserve Board Working Paper, FEDS 2026018.

DeepSeek-AI, A. Liu, B. Feng, B. Xue, et al. (2024) "DeepSeek-V3 Technical Report," arXiv Preprint. Available at: https://arxiv.org/abs/2412.19437.

Dell, M. (2025) "Deep Learning for Economists," *Journal of Economic Literature*, Vol. 63, No. 1, pp. 5–58.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024) "GPTs are GPTs: Labor Market Impact Potential of LLMs," *Science*, Vol. 384, No. 6702, pp. 1306–1308.

Felten, E.W., M. Raj, and R. Seamans (2019) "The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization," Working Paper, NYU Stern School of Business.

Feng, D., X. Wei, L. Rosenbaum, A. Maki, and K. Dietmayer (2019) "Deep Active Learning for Efficient Training of a Lidar 3d Object Detector," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 667–674.

Gentzkow, M., B. Kelly, and M. Taddy (2019) "Text as Data," *Journal of Economic Literature*, Vol. 57, No. 3, pp. 535–574.

Gerstgrasser, M., R. Schaeffer, A. Dey, R. Rafailov, H. Sleight, J. Hughes, T. Korbak, R. Agrawal, D. Pai, A. Gromov, D.A. Roberts, D. Yang, D.L. Donoho, and S. Koyejo (2024) "Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data," arXiv Preprint arXiv:2404.01413.

Goldfarb, A., B. Taska, and F. Teodoridis (2023) "Could Machine Learning be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings," *Research Policy*, Vol. 52, No. 1, article 104653.

Gu, A. and T. Dao (2023) "Mamba: Linear-time Sequence Modeling with Selective State Spaces," arXiv Preprint arXiv:2312.00752.

Handa, K., A. Tamkin, M. McCain, S. Huang, E. Durmus, S. Heck, J. Mueller, J. Hong, S. Ritchie, T. Belonax, K.K. Troy, D. Amodei, J. Kaplan, J. Clark, and D. Ganguli (2025) "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," arXiv Preprint arXiv:2503.04761.

Hart, G.L.W., T. Mueller, C. Toher, and S. Curtarolo (2021) "Machine Learning for Alloys," *Nature Reviews Materials*, Vol. 6, No. 8, pp. 730–755.

Hinton, G., O. Vinyals, and J. Dean (2015) "Distilling the Knowledge in a Neural Network," arXiv Preprint arXiv:1503.02531. Available at: https://arxiv.org/abs/1503.02531.

Ho, A., T. Besiroglu, E. Erdil, D. Owen, R. Rahman, Z.C. Guo, D. Atkinson, N. Thompson, and J. Sevilla (2024) "Algorithmic Progress in Language Models," arXiv Preprint arXiv:2403.05812.

Iansiti, M. and K.R. Lakhani (2020) *Competing in the Age of AI: Strategy and Leadership when Algorithms and Networks run the World*, Harvard Business Press, Cambridge, Massachusetts.

Jacobs, R.A., M.I. Jordan, S.J. Nowlan, and G.E. Hinton (1991) "Adaptive Mixtures of Local Experts," *Neural Computation*, Vol. 3, No. 1, pp. 79–87.

Jiang, A.Q., A. Sablayrolles, A. Mensch, C. Bamford, D.S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L.R. Lavaud, M.-A. Lachaux, P. Stock, T. Le Scao, T. Lavril, T. Wang, T. Lacroix, and W. El Sayed (2023) "Mistral 7B," arXiv Preprint arXiv:2310.06825.

Johnson, S. (2024) "NotebookLM Gets a New Look, Audio Interactivity and a Premium Version," *Google: The Keyword*, December 13, available at: https://blog.google/innovation-and-ai/models-and-research/google-labs/notebooklm-new-features-december-2024/.

Kalidindi, S.R., M. Buzzy, B.L. Boyce, and R. Dingreville (2022) "Digital Twins for Materials," *Frontiers in Materials*, Vol. 9, article 818535.

Kane, A. and M.N. Baily (2025a) "AI in the Electricity Sector: Optimizing Grid Management and Energy Use," Brookings Institution.

Kane, A. and M.N. Baily (2025b) "AI in the Information Sector: Advancing Software, Customer Service, and Design," Brookings Institution.

Kaplan, J., S. McCandlish, T. Henighan, T.B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei (2020) "Scaling Laws for Neural Language Models," arXiv Preprint arXiv:2001.08361.

Korinek, A. (2023) "Generative AI for Economic Research: Use Cases and Implications for Economists," *Journal of Economic Literature*, Vol. 61, No. 4, pp. 1281–1317.

Krizhevsky, A., I. Sutskever, and G.E. Hinton (2012) "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, Vol. 25.

Li, K., A.K. Hopkins, D. Bau, F. Viégas, H. Pfister, and M. Wattenberg (2022) "Emergent World Representations: Exploring a Sequence Model Trained on a Synthetic Task," arXiv Preprint arXiv:2210.13382.

Lipsey, R.G., K.I. Carlaw, and C.T. Bekar (2005) *Economic Transformations: General Purpose Technologies and Long-term Economic Growth*, OUP Oxford.

Liu, G., F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro (2018) "Image Inpainting for Irregular Holes using Partial Convolutions," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 85–100.

Liu, R., J. Wei, F. Liu, C. Si, Y. Zhang, J. Rao, S. Zheng, D. Peng, D. Yang, D. Zhou, and A.M. Dai (2024) "Best Practices and Lessons Learned on Synthetic Data for Language Models," arXiv Preprint arXiv:2404.07503.

Lugmayr, A., M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool (2022) "Repaint: Inpainting using Denoising Diffusion Probabilistic Models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471.

McKinsey (2024) "The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value," McKinsey & Company.

Mokyr, J. (2004) *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton University Press.

Mowery, D.C. and N. Rosenberg (1999) *Paths of Innovation: Technological Change in 20th-Century America*, Cambridge University Press, Cambridge, England.

Nordhaus, W.D. (2007) "Two Centuries of Productivity Growth in Computing," *The Journal of Economic History*, Vol. 67, No. 1, pp. 128–159.

OpenAI (2023) "Introducing GPTs," November 6, available at: https://openai.com/index/introducing-gpts/.

Ouyang, L., J. Wu, X. Jiang, D. Almeida, C.L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe (2022) "Training Language Models to Follow Instructions with Human Feedback," *Advances in Neural Information Processing Systems*, Vol. 35, pp. 27730–27744.

Pairolero, N.A., A.V. Giczy, G. Torres, T.I. Erana, M.A. Finlayson, and A.A. Toole (2025) "The Artificial Intelligence Patent Dataset (AIPD) 2023 Update," *The Journal of Technology Transfer*, pp. 1–24.

Park, J.S., J. O'Brien, C.J. Cai, M.R. Morris, P. Liang, and M.S. Bernstein (2023) "Generative Agents: Interactive Simulacra of Human Behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22.

Pope, R., S. Douglas, A. Chowdhery, J. Devlin, J. Bradbury, J. Heek, K. Xiao, S. Agrawal, and J. Dean (2023) "Efficiently Scaling Transformer Inference," *Proceedings of Machine Learning and Systems*, Vol. 5, pp. 606–624.

Porter, M.E. and S. Stern (2001) "Innovation: Location Matters," *MIT Sloan Management Review*.

Reed, S., K. Zolna, E. Parisotto, S.G. Colmenarejo, A. Novikov, G. Barth-Maron, M. Gimenez, Y. Sulsky, J. Kay, J.T. Springenberg, T. Eccles, J. Bruce, A. Razavi, A. Edwards, N. Heess, Y. Chen, R. Hadsell, O. Vinyals, M. Bordbar, and N. de Freitas (2022) "A Generalist Agent," arXiv arXiv:2205.06175.

Saadi, J.I. and M.C. Yang (2023) "Generative Design: Reframing the Role of the Designer in Early-Stage Design Process," *Journal of Mechanical Design*, Vol. 145, No. 4, article 041411.

Serradilla, O., E. Zugasti, J. Rodriguez, and U. Zurutuza (2022) "Deep Learning Models for Predictive Maintenance: A Survey, Comparison, Challenges and Prospects," *Applied Intelligence*, Vol. 52, No. 10, pp. 10934–10964.

Shazeer, N., A. Mirhoseini, K. Maziarz, A. Davis, Q.V. Le, G.E. Hinton, and J. Dean (2017) "Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer," in *Proceedings of the 5th International Conference on Learning Representations (ICLR)*. Available at: https://arxiv.org/abs/1701.06538.

Shree, P. (2020) "The Journey of Open AI GPT Models," *Walmart Global Tech Blog*, November 9, available at: https://medium.com/walmartglobaltech/the-journey-of-open-ai-gpt-models-32d95b7b7fb2.

Shumailov, I., Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, and R. Anderson (2023) "The Curse of Recursion: Training on Generated Data Makes Models Forget," arXiv Preprint arXiv:2305.17493.

Tao, F., H. Zhang, and C. Zhang (2024) "Advancements and Challenges of Digital Twins in Industry," *Nature Computational Science*, Vol. 4, No. 3, pp. 169–177.

Taori, R., I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T.B. Hashimoto (2023) "Stanford Alpaca: An Instruction-following Llama Model," Stanford Center for Research on Foundation Models.

Trajtenberg, M. (2018) "AI as the Next GPT: A Political-Economy Perspective," NBER Working Paper 24245, National Bureau of Economic Research.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin (2017) "Attention is All You Need," *Advances in Neural Information Processing Systems*, Vol. 30.

Wang, D. and A.-L. Barabási (2021) *The Science of Science*, Cambridge University Press.

Wang, H., S. Ma, L. Dong, S. Huang, H. Wang, L. Ma, F. Yang, R. Wang, Y. Wu, and F. Wei (2023) "BitNet: Scaling 1-Bit Transformers for Large Language Models," arXiv Preprint arXiv:2310.11453.

Webb, M. (2019) "The Impact of Artificial Intelligence on the Labor Market," Available at SSRN 3482150.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou (2022) "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," *Advances in Neural Information Processing Systems*, Vol. 35, pp. 24824–24837.

Whitehead, A.N. (1925) *Science and the Modern World: Lowell Lectures, 1925*, New American Library.

Yin, M., H. Vu, and C. Persico (2026) "How (un)Stable Are LLM Occupational Exposure Scores? Evidence from Multi-Model Replication," NBER Working Paper w35110.

Zeff, M. (2024) "Current AI Scaling Laws Are Showing Diminishing Returns, Forcing AI Labs to Change Course," *TechCrunch*, November 20, available at: https://techcrunch.com/2024/11/20/ai-scaling-laws-are-showing-diminishing-returns-forcing-ai-labs-to-change-course/.

The Productivity J-Curve from an International Perspective: Is the United States Unique?

Ahmed Bounfour, Kazuma Edamura, Takayuki Ishikawa,
Tsutomu Miyagawa, Alberto Nonnis and Konomi Tonogi*

IPM Research Article

Abstract

Despite major advances in technology, productivity growth has slowed across advanced economies. Brynjolfsson, Rock and Syverson (2021) argued that, due to the large intangible investments associated with digitalization, total factor productivity (TFP) growth may be underestimated and be higher after the investment boom. The resulting gap between standard and revised measures produces a J-shaped pattern, the "productivity J-curve".

Building on this work, we examine productivity estimates for five advanced economies: France, Germany, Japan, the United Kingdom and the United States. Using the estimated coefficients on intangibles (research and development, software, and organizational capital) in the value functions of listed firms over 2006-2020, we find that TFP underestimation caused by large intangible investments was largely unique to the United States, and was much smaller in Europe and Japan. This result is consistent with the recent productivity rebound in the U.S. and reinforces the call for investment in innovation in Europe and Japan.

*Authors are listed in alphabetical order and all contributed equally to this article. Ahmed Bounfour: Université Paris-Saclay and Africa Business School/UM6P is the corresponding author (Email: ahmed.bounfour@universite-paris-saclay.fr). Kazuma Edamura: Kanagawa University. Takayuki Ishikawa: Kanagawa University. Tsutomu Miyagawa: RIETI and Gakushuin University. Alberto Nonnis: ESLSCA Business School Paris. Konomi Tonogi: Rissho University. We thank anonymous referees for constructive comments. We thank Professors Abdul Erumban, Cecilia Jona-Lasinio, Kenji Tanaka and Bart van Ark for their insightful suggestions and comments at Gakushuin University, the 8th World KLEMS Conference, the Japan Economic Association 2025 Spring meeting and the 10th Society for the Economic Measurement conference. This study is conducted as a part of the Project "On Productivity Growth through Comprehensive Capital Accumulation" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). We would like to thank Dr. Kyoji Fukao, Dr. Eiichi Tomiura, and Professor YounGak Kim of the RIETI DP Seminar for their helpful comments. Our study is supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (No. 22KK0021 and 23K01381) of Japan, and the Japan Securities Scholarship Foundation. Support was also provided by the European chair on intangibles of the Université Paris-Saclay, for the French team. Any remaining errors should be attributed to the authors.

1. Introduction

Digitalization in the 21st century has revolutionized our lives and businesses. While one would expect this profound digital transformation to lead to faster productivity growth and higher living standards, productivity growth rates in the advanced countries have generally stagnated since the 2010s (Fernald, Inklaar and Ruzic (2025), Goldin, Koutroumpis, Lafond and Winkler (2024) and Van Ark, de Vries and Erumban (2024)). Over this time, two major digital innovations have emerged: first, the rise of platform organizations such as Airbnb and Uber in the early 2010s, and second, the development of generative artificial intelligence (AI) businesses such as OpenAI. These seemingly conflicting developments — of faster digitalization and slower productivity growth — have led to academic and policy debates on the economic effects of digitalization on productivity growth.

The issues can be framed within the broader challenge of measuring the impact of the digital economy, or any other technological revolution, on growth. As argued by Aghion *et al.* (2019), official statistics do not correctly capture prices. Because new entrants tend to offer lower prices than incumbents, true price levels are lower than those published by official statistics, and the true real value added is higher.

One specific challenge concerns the accurate measurement of intangible capital in the national accounts framework. Following the standard measure, total factor productivity (TFP) growth is low during periods of high investment in digitalization, because increasing adjustment costs associated with new investments depress gross domestic product (GDP). As a result, TFP growth decreases during the investment boom.

Brynjolfsson, Rock and Syverson (2021) developed an alternative, finance-based approach to measuring intangibles. They build on the standard neoclassical investment theory with adjustment costs — as developed by Lucas (1967) and Uzawa (1969) — which holds that capital formation requires additional expenditures (such as employee training and organizational change) which should be treated as capital investment, even though traditional investment treats them as temporary outlays. They then estimated the parameters of adjustment costs of investment and used them to revise the standard measure of TFP growth.

Once the adjustment costs are recognized as intangible investments, the revised GDP does not fall and the revised TFP growth rate becomes stable. Because the standard TFP growth rate recovers after the investment boom, the movements in the gap between the standard TFP growth rate and the revised TFP growth rate resemble the letter J. Hence BRS (2021) called these movements the "productivity J-curve".[1] They argued that this model can explain the low productivity rate during the investment boom in the platform

1 According to BRS (2021), Larry Summers pointed out that the shape of productivity resembles the J-curve in the balance of trade literature.

industry and AI. Miyagawa, Tonogi and Ishikawa (2021) found that, in Japan, the difference between standard and revised TFP in information and communications technology-intensive industries followed a J-curve shape at least twice between the late 1990s and the 2010s.

In this paper, we extend the approach to measure intangibles and the revised TFP growth rate developed by BRS (2021) to several large countries in Europe as well as the U.S. and Japan. As Goldin *et al.* (2024) pointed out, not only Japan and the U.S., but also many European countries suffered from a productivity slowdown in the official statistics in the 2010s. However, the speed of digitalization and technological progress in the U.S. may have been faster than in the other advanced countries. For example, due to strict regulations, ride-sharing services have not yet been permitted in Japan, and the EU has regulated the monopolistic behavior by gatekeepers such as Amazon and Google.

The aim of our article is twofold. First, we examine to what extent the productivity slowdown in the advanced countries in the 2010s is overstated, by computing TFP measures adjusted for intangible capital, and examining whether the resulting gap follows a productivity J-curve pattern. This allows us to evaluate whether TFP growth during the 2010s is likely to have been underestimated. Second, we aim to clarify the scale and speed of digitalization by measuring underestimation of TFP growth induced by software investment.

Our empirical findings show that the productivity slowdown observed in the 2010s may have been overstated, as TFP was underestimated to some extent in all advanced countries. When focusing on capital formation such as R&D, software and organizational capital required to introduce new technologies, we find that productivity growth in the U.S. is largely underestimated, consistently with what was found by BRS (2021). In particular, rapid accumulation of intangible assets associated with software investment generated underestimation of U.S. TFP growth. Compared to the U.S., the scale of the underestimation of TFP growth rates in European countries and Japan is relatively small. Our results imply that European and Japanese firms depend on software created by U.S. firms which make large outlays to develop software. The low associated costs with capital formation in organizational capital in European countries and Japan imply that firms have been reluctant to make large investments in organizational change — which would have generated large adjustment costs, as was the case in the United States.

This article consists of six sections. In the next section, we review the literature on measurement issues related to GDP and intangibles, the productivity slowdown in advanced countries, and capital formation with multiple assets. In the third section, we present our data and model to estimate the parameters of adjustment costs of investment. Using listed firms' data in five advanced countries (France, Germany, Japan, the UK and the U.S.), we estimate two types of equations where explanatory variables consist of multiple assets: one consists of four assets (construction and buildings, machinery, R&D and software) and the other of five assets by adding organizational capital.

The fourth section shows our estimation results. We find that many assets are accompanied by associated costs with investment. In particular, investment in R&D assets is accompanied by high costs, for example because of the need for highly paid, high-skilled R&D workers. In the case of the U.S., the estimated coefficients for all assets suggest that capital formations generate large expenditures for intangibles which are consistent with BRS (2021).

The fifth section computes the revised TFP growth rates, using the parameters from these estimations, and examines whether the resulting gap relative to the standard TFP measures produces under- and over-estimations compatible with the J-curve theory. First, we revise TFP to account for intangibles associated with R&D, software and organizational capital, which are the key components of recent technological progress. As for R&D, we find that the J-curve effects generated by this type of asset are rather small, indicating minimal measurement issues associated with R&D. For software, we find that the U.S. incurred substantial associated costs as rapid digitalization accelerated software investment during the 2010s. Our results suggest that the U.S. productivity growth rate is underestimated by up to 1.4 percentage points when correcting for software, and 1.6 when correcting for software, R&D, and organizational capital. In France and Japan, we find small underestimation of TFP growth rate generated by the organizational capital, implying that those firms appear to have avoided such costs by limiting organizational investment. When we correct TFP for all intangible types in our data collectively, using the combined estimated coefficients, we again find a large underestimation of TFP growth in the United States.

In the last section, we summarize implications obtained from our estimation results and policy implications from our study.

2. Related Literature

Our paper relates to multiple research areas, such as measurement issues on GDP and intangibles, the productivity slowdown in advanced countries, and capital formation with multiple assets. Regarding the measurement of GDP, Erik Brynjolfsson and co-authors published at least three relevant articles. For example, Brynjolfsson *et al.* (2019) discuss the measurement of GDP noting that, as official GDP is a measure from the production side, the value of software provided for free is not counted in the current GDP framework. Hence, they suggest that GDP in the digital age should be measured from the consumer side, which they call "GDP-B". Tambe *et al.* (2020) measure digital capital based on the number of workers in digital firms, and find a positive relationship between digital capital and productivity growth. Brynjolfsson, Rock and Syverson (2021, henceforth BRS), which focuses on the measurement of intangibles and TFP growth, was discussed in the introduction as the central reference for our study.[2]

2 Basu *et al.* (2003) suggested that adjustment costs in the neoclassical theory should be recognized as intangibles.

Coyle and Nakamura (2022) also argue that GDP should be measured, not from the production side, but the consumer side. Hasegawa (2023) and Miyagawa (2024) refer to the experimental measures of the digital economy by the Cabinet Office of the Government of Japan which show that the digital economy accounted for 8.6 per cent of GDP in Japan in 2018. Miyagawa (2024) also measured the scale of digitalized inputs that are not recognized as assets in the current System of National Accounts (SNA), such as the cloud and AI procured services. Finally, the World Bank estimates that the digital economy contributes more than 15 per cent to global domestic product (Hayat, 2022).

There are many studies on the productivity slowdown in advanced countries. As stated in the introduction, Aghion *et al.* (2019) argue that true productivity would be higher than official productivity growth, if the statistics correctly captured the new economy due to digitalization. In contrast, Gordon (2016) argues that the recent U.S. productivity slowdown is to be expected because recent digital transformation has been less effective in improving living standards compared to the main innovations of the 20th century, such as running water, electricity, automobiles and washing machines. Acemoglu *et al.* (2014) argues that slower U.S. productivity growth was caused mainly by the decline in manufacturing employment, rather than technological progress as a result of digitalization. Goldin *et al.* (2024) examine which factors affected the productivity slowdown in the advanced countries by decomposing labour productivity growth into the contributions from capital deepening, labour composition, and TFP growth. In the U.S., both capital deepening effects and TFP growth slowed labour productivity growth, while slower TFP growth was the major factor behind the slowdown in France. Mismeasurement and allocative inefficiency had particularly large effects on French TFP growth. In Japan, capital deepening has slowed significantly while spillover effects from intangibles and trade are also found to be a crucial factor for weaker productivity growth.

The theoretical background of our paper is based on the neoclassical theory of investment with multiple assets. Following the theory of investment with a single asset (Lucas, 1967; Uzawa, 1969; Hayashi, 1982), the theory with multiple assets was developed by Wildasin (1984). Hall (2001), Miyagawa and Kim (2008) and Miyagawa, Takizawa and Edamura (2015) show that a firm's value can be expressed as the weighted sum of each asset under the assumption of linear homogeneous production and investment functions. Hall (2001) argues that when Tobin's q for a firm exceeds 1 it indicates the value of intangibles. This indicates that the stock market evaluates unmeasured intangibles in a firm's balance sheet. Following this argument, Miyagawa *et al.* (2015) show that Tobin's q, when only measured on the basis of tangible capital, is greater than 1 for the Japanese ICT firms. However, when considering unmeasured intangibles on these firms' balance sheets, the revised Tobin's q becomes closer to 1. This suggests that unmeasured intangibles in ICT firms contribute to increased firm valuations.

3. Data

As noted by BRS (2021) and further developed by Miyagawa, Tonogi and Ishikawa (2021) and Bijnens, Konings and Putseys (2025), the introduction of a new General-Purpose Technology (GPT) often triggers an initial phase of investment in many associated intangible assets which may go unaccounted for, resulting in TFP mismeasurement. This measurement problem could arise from excluding certain intangible assets from national accounts, which BRS (2021) call "intangible correlates". These complementary, correlated intangible investments — such as those used to adapt workers and business organizational structures to technological advancements — often do not appear in national accounts because they lack physical form, have ad hoc characteristics, are difficult to quantify, and are often treated as expenses rather than capitalized as investment (Bavdaž *et al.*, 2023; Bounfour and Nonnis, 2025; Nonnis, Bounfour and Kim, 2023). As pointed out by Brynjolfsson, Hitt and Yang (2002) and BRS (2021), the stock market valuation of firms differs from book values at least partly due to the valuation of these intangible correlates, which are not visible in the books, but are reflected in the firm's overall market capitalization.

Assuming financial markets correctly evaluate these intangible assets, their implicit worth can be retrieved by estimating "market value regressions", where the market value of the firm is regressed on its tangible and intangible assets:

$$V_{it} = const. + a_1 TA_{1it-1} + a_2 TA_{2it-1} + a_3 RD_{it-1} + a_4 SOFT_{it-1} + \mu_t + v_i + \epsilon_{it} \tag{1}$$

$$V_{it} = const. + b_1 TA_{1it-1} + b_2 TA_{2it-1} + b_3 RD_{it-1} + b_4 SOFT_{it-1} + b_5 ORG_{it-1} + \mu_t + v_i + \epsilon_{it} \tag{2}$$

In equations (1) and (2), V_{it} is the firm value of firm i at time t, with $V_{it} = p_{sit} S_{it} + D_{it}$, where p_{sit} is the share price, S_{it} is number of shares outstanding and D_{it} is debt for firm i. TA_{1it} and TA_{2it} are the tangible assets of buildings and construction, and machinery. RD_{it}, $SOFT_{it}$, and ORG_{it}, are the research and development capital stock, software assets, and organizational capital, respectively.

We obtain all data, except for $SOFT_{it}$ and ORG_{it}, directly from the Orbis dataset. The software variable $SOFT_{it}$ for each firm i is calculated by multiplying the total assets of firm i by the ratio of software to total assets at the industry level, obtained from the EUKLEMS/INTANProd data released in 2023 (Bontadini *et al.*, 2023; 2024) and the Japanese Industrial Productivity (JIP) 2023 database.

Some software programs such as AI and online meeting tools are often on a subscription basis, whose costs are not counted as assets but as part of sales, general and administration costs (SG&A). The Basic Survey of Japanese Business Structure and Activities (BSBSA) conducted by Ministry of Economy, Trade and Industry shows that the share of these ICT costs in the total

SG&A costs is 3 per cent. Using this data, and a depreciation rate of 33 per cent based on the Japanese SNA, we capitalize information and communication costs in SG&A. Then, $SOFT_{it}$ is a sum of software stock constructed from the industry-level data (EUKLEMS/INTANProd Data), and capitalized assets are constructed from information and communication costs data (BSBSA survey data).[3]

Although we use the SG&A data from the Orbis dataset, we make additional manipulations to construct ORG.[4] Hulten and Hao (2008) and Eisfeldt and Papanikolaou (2013) recognized 30 per cent of SG&A costs as capital formation in organizational capital. Therefore, while we recognize one-tenth of the organizational capital defined in previous studies as software investment, we classify the remainder as investment in our newly defined organizational capital. We construct organizational capital stock by the perpetual inventory method, with a depreciation rate of 40 per cent based on Corrado, Hulten and Sichel (2009).

Table A1 in the annex shows summary statistics for firms in France, Germany, Japan, the UK and the United States. We expect all coefficients on each asset to be positive. When the coefficient of an asset is greater than its asset price, this shows that capital formation in this asset is accompanied by adjustment costs that are accumulated as intangibles.[5]

4. Estimating Firm Value Functions

We estimated equations (1) and (2) for the period from 2006 to 2020 using both pooled OLS and system GMM estimators (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998), which allowed us to control for the potential endogeneity of the explanatory variables. In the GMM specification, we treat firms' investment decisions as fully endogenous and use lagged variables starting from at least the second lag as instruments. This assumes that capital assets may be correlated with current and past shocks, but uncorrelated with future shocks (Ackerberg, Caves and Frazer, 2015).[6] The lag structure in each estimation is selected in order to satisfy the Hansen test for overidentifying restrictions and the Arellano-Bond AR(2) test for second order serial correlation in the residuals.[7]

3 Although BSBSA covers only Japanese firms, we use the ratio of information and communication costs in the total SG&A costs in all samples, because these subscription costs are not counted as assets in firms in the advanced countries.

4 SG&A is widely used as a proxy for organizational capital measurement in finance literature (see Lev and Radhakrishnan, 2003).

5 However, as price indices usually move around 1, we focus on whether an estimated coefficient is greater than 1 for the condition to draw a productivity J-curve.

6 Our endogeneity assumption is relatively strong, as it allows regressors to be correlated with current shocks. A less restrictive assumption would treat regressors as predetermined, i.e. uncorrelated with current shocks, and use lagged variables starting from the first lag as instruments, but resulted in unsatisfactory diagnostics (Hansen and AR(2) tests).

7 We select estimations in which the Hansen test p-value is significant (above 0.10) but not too large, to avoid the risk of instrument proliferation, ideally targeting values around 0.2.

The estimation results for Equations (1) and (2) are reported in Tables 1 and 2, respectively. In each table, the model is estimated for the full sample (columns 1-2) and separately for each country (columns 3-12). Odd-numbered columns are pooled OLS estimates, while even-numbered columns are system GMM estimates.

Tables 1-1 and 1-2 show that almost all coefficients are positive and significant. This implies that all assets contribute positively to firm valuations. In the estimations using data from all countries, all OLS coefficients are positive, significant and greater than 1, even though the coefficient of machinery assets is less than 1 in the GMM estimation. The coefficients for intangibles (R&D and software) are greater than 1 in both estimations.

The country-level estimation results differ across countries. The results for the U.S. (columns 11 and 12) indicate that all assets contribute to the accumulation of intangibles, as their coefficients are positive, significant and greater than 1. These results are consistent with those for the U.S. shown in BRS (2021). In the other countries, the majority of R&D and software coefficients are greater than 1, indicating positive hidden adjustment costs or correlated unmeasured intangibles. Following BRS (2021), we interpret these coefficients as reflecting financial markets' valuation of intangible assets relative to their theoretical investment costs. Occasionally, country-specific coefficients fall below one (e.g., R&D in Columns 5 and 10, or software in columns 3, 4 and 7).[8] In principle, this would imply that financial markets value the asset less than its theoretical investment cost, suggesting rapid obsolescence or inefficiency. However, given the inherent disruptive nature of these assets and the prevalence of coefficients greater than one in our estimations, we interpret these lower values as anomalies, due to statistical limitations of the econometric model or financial market inefficiencies. Therefore, rather than assuming a theoretically implausible negative intangible investment, we substitute these specific estimates with their corresponding full sample baseline to maintain economic consistency.[9]

The same rationale is applied to the estimation of Equation (2) in Table 2. While the number of positive and significant coefficients greater than 1 is lower than in Table 1, the U.S. estimates remain robust, with all intangible coefficients positive, significant and greater than 1. All coefficients of software capital in Table 2 are smaller than those in Table 1. Especially, software coefficients in Japan are less than 1. As for organizational capital, we find positive and significant coefficients greater than 1, except in column 10.

8 The low costs associated with software investment in Japan may be due to the large share of customized software in Japan, which does not require additional training costs for employees.

9 As a robustness check, an alternative approach would be to constrain these coefficients to 1, assuming therefore zero unmeasured intangible investment rather than negative. This test does not alter our final conclusions in the next section.

Table 1-1: Estimation Results of Equation (1), from All Samples to Germany

Four asset cases

	(1) All pooled	(2) All GMM	(3) France pooled	(4) France GMM	(5) Germany pooled	(6) Germany GMM
Buildings and construction	2.09***	2.67***	1.21***	1.91*	2.08***	2.92
	(0.04)	(0.54)	(0.20)	(0.99)	(0.13)	(1.85)
Machinery	1.24***	0.60***	1.69***	2.10**	1.03***	0.71***
	(0.02)	(0.28)	(0.11)	(1.02)	(0.05)	(0.21)
Research and development	3.02***	3.56***	5.86***	8.53***	0.60***	2.84**
	(0.05)	(0.76)	(0.57)	(3.38)	(0.14)	(1.34)
Software	1.07***	1.28***	0.97***	0.60**	9.68***	5.99***
	(0.03)	(0.47)	(0.08)	(0.26)	(0.46)	(3.02)
Constant	103.31	-95.26	2,529.95	-255.40	167.97	-120,036.00
	(3,184.88)	(307.29)	(5,725.32)	(4,218.69)	(5,405.62)	(285,800.00)
Observations	29,792	29,792	2,057	2,057	2,058	2,058
Number of groups		2,345		172		158
Adjusted R-squared	0.56	0.41	0.62	0.27	0.77	0.86
Hansen test (p value)		0.12		0.23		0.50
AR (2) test (p value)		0.41		0.27		0.86

Note: Standard errors shown below estimated coefficients

Source: Authors' calculations. • Created with Datawrapper

5. Revised TFP and Productivity J-Curves

Following BRS (2021), we revised TFP measures to account for intangible correlates using the estimated coefficients in Equations (1) and (2). As explained in Section 1, estimated coefficients include adjustment costs of investment. If an estimated coefficient of an asset i divided by price of asset i is higher than 1, we are able to identify intangible correlates associated with capital formation in asset i and revise the standard measure of TFP growth rate accordingly. The gap between the standard measure of TFP growth, g_A, and the adjusted measure, g_A^*, can be expressed as follows:

$$g_A - g_A^* = \theta g_A - \theta(g_{IZ} - g_K) \qquad (3)$$

Table 1-2: Estimation Results of Equation (1), from Japan to U.S.

Four asset cases

	(7) Japan pooled	(8) Japan GMM	(9) UK pooled	(10) UK GMM	(11) U.S. pooled	(12) U.S. GMM
Buildings and construction	2.55***	2.55**	0.49*	4.33*	2.08***	2.25***
	(0.04)	(0.99)	(0.27)	(2.27)	(0.06)	(0.73)
Machinery	0.05**	0.180	3.03***	1.240	1.32***	1.14***
	(0.02)	(0.65)	(0.18)	(1.77)	(0.04)	(0.46)
Research and development	2.21***	2.32***	1.71***	0.610	6.28***	6.19***
	(0.04)	(0.68)	(0.15)	(1.03)	(0.13)	(0.71)
Software	0.59***	1.23**	4.55***	4.03**	6.99***	7.87***
	(0.02)	(0.50)	(0.21)	(1.43)	(0.23)	(2.99)
Constant	-8.83	393.50***	787.39	745.20	190.36	68.81
	(2,316.49)	(182.18)	(2,271.90)	(543.94)	(5,558.69)	(208.52)
Observations	14,650	14,650	2,432	2,432	7,734	7,734
Number of groups		1,161		207		566
Adjusted R-squared	0.75	0.09	0.58	0.46	0.67	0.32
Hansen test (p value)		0.17		0.15		0.20
AR (2) test (p value)		0.09		0.46		0.32

Note: ***, **, and * indicate significance at 1, 5, and 10 per cent levels, respectively. Standard errors are reported in parentheses. The table reports estimation results from equation (1), where the dependent variable is firm market value and the explanatory variables are capital stock assets. All regressions are estimated using pooled OLS (odd-numbered columns) and system GMM (even-numbered columns). Regressions include industry, year and industry-year interaction dummy variables. In the AR tests, the null hypothesis is the absence of serial correlation in the error term. In the Hansen test, the null hypothesis is the exogeneity of instruments. Instruments are chosen to maximize the efficiency of the Hansen test. For each country, observations with values of the firm value exceeding ±3 standard deviations from the mean have been removed as outliers.

Source: Authors' calculations. • Created with Datawrapper

The gap evidenced in Equation (3) often results in a J-shaped curve (productivity J-curve), which is typical of the investment cycle of disruptive technologies. In the equation, θ is the share of intangible investment in value added, when including intangible investment. When intangible investment, (g_{IZ}), measured by including the associated costs of capital formation, exceeds tangible capital growth, (g_K), the gap between the standard and revised TFP growth becomes negative, meaning that the standard TFP growth rate is underestimated, due to the incorrect measurement of intangible capital.

We construct four revised TFP measures both by accounting for each intangible capital type individually, and all three (R&D, software, and organizational capital) together. To do so, we use the GMM

Table 2-1: Estimation Results of Equation (2), from All Samples to Germany

Five assets cases

	(1) All pooled	(2) All GMM	(3) France pooled	(4) France GMM	(5) Germany pooled	(6) Germany GMM
Buildings and construction	1.19***	0.95	0.55***	-0.33	0.39***	-0.86
	(0.04)	(1.03)	(0.19)	(1.41)	(0.12)	(1.86)
Machinery	0.91***	0.35	0.42***	1.11	0.90***	0.65**
	(0.02)	(0.30)	(0.13)	(0.87)	(0.04)	(0.30)
Research and development	2.11***	2.20**	7.00***	5.86**	0.39***	0.01
	(0.05)	(1.11)	(0.54)	(2.44)	(0.11)	(1.29)
Software	0.74***	0.73*	0.67***	0.53***	2.42***	2.81
	(0.03)	(0.38)	(0.08)	(0.19)	(0.43)	(2.40)
Organizational capital	2.74***	3.80*	1.73***	1.96*	4.08***	4.79***
	(0.04)	(2.09)	(0.11)	(1.03)	(0.13)	(1.81)
Constant	68.22	99.80	5,231.61	364.20	-15.31	-41,727.70
	(2,988.90)	(311.88)	(5,390.60)	(319.47)	(4,287.70)	(80,244.23)
Observations	29,792	29,792	2,057	2,057	2,058	2,058
Number of groups		2,345		172		158
Adjusted R-squared	0.61	0.41	0.66	0.21	0.85	1.00
Hansen test (p value)		0.19		0.15		0.13
AR (2) test (p value)		0.41		0.21		1.00

Note: Standard errors shown below estimated coefficients

Source: Authors' calculations. • Created with Datawrapper

coefficients from Tables 1 and 2 if the country-level coefficients are positive, significant, and larger than the respective asset prices. If a coefficient does not meet these conditions, we replace it with the corresponding full-sample coefficient from columns 1 and 2 in Tables 1 and 2, which pools all countries together. The estimated coefficients used to measure revised TFP growth rate are listed in Table 3, where

10 In Table 3, we use estimated coefficients from GMM estimations in Table 1 for R&D and software and estimated coefficients from GMM estimations in Table 2 for organizational capital. We do so to avoid using coefficients that are not positive, significant, or larger than the respective asset prices. However, using different and more consistent rules did not alter much our results and we do not report them in the main paper. We also measure productivity J-curves using the pooled estimation results. In this case, we also have the same results as the case using the GMM estimation results. When we estimate Equation (1) and (2) including outliers, some estimations do not satisfy the Hansen test. However, productivity J-curves using parameters from pooled

Table 2-2: Estimation Results of Equation (2), from Japan to U.S.

Five assets cases

	(7) Japan pooled	(8) Japan GMM	(9) UK pooled	(10) UK GMM	(11) U.S. pooled	(12) U.S. GMM
Buildings and construction	1.70***	2.39***	-1.19***	5.17**	0.76***	-0.41
	(0.04)	(0.90)	(0.27)	(2.39)	(0.07)	(1.04)
Machinery	-0.02	-0.46	2.74***	0.63	1.02***	0.60**
	(0.02)	(0.40)	(0.17)	(1.44)	(0.04)	(0.30)
Research and development	0.32***	1.26**	-0.37**	0.41	5.16***	5.47***
	(0.05)	(0.61)	(0.17)	(1.48)	(0.12)	(1.04)
Software	0.36***	0.67***	1.37***	4.36***	4.75***	3.79**
	(0.02)	(0.26)	(0.26)	(1.62)	(0.22)	(1.61)
Organizational capital	3.90***	3.08***	4.58***	-0.01	4.47***	5.74***
	(0.07)	(1.13)	(0.24)	(0.24)	(0.14)	(2.17)
Constant	-99.55	237.60	783.82	1077.80**	175.50	314.40
	(2,116.73)	(149.43)	(2,095.87)	(677.86)	(5,216.76)	(247.60)
Observations	14,650	14,650	2,432	2,432	7,734	7,734
Number of groups		1,161		207		566
Adjusted R-squared	0.79	0.15	0.64	0.42	0.71	0.23
Hansen test (p value)		0.17		0.11		0.26
AR (2) test (p value)		0.15		0.42		0.23

Note: ***, **, and * indicate significance at 1, 5, and 10 per cent levels, respectively. Standard errors are reported in parentheses. The table reports estimation results from equation (1), where the dependent variable is firm market value and the explanatory variables are capital stock assets. All regressions are estimated using pooled OLS (odd-numbered columns) and system GMM (even-numbered columns). Regressions include industry, year and industry-year interaction dummy variables. In the AR tests, the null hypothesis is the absence of serial correlation in the error term. In the Hansen test, the null hypothesis is the exogeneity of instruments. Instruments are chosen to maximize the efficiency of the Hansen test. For each country, observations with values of the firm value exceeding ±3 standard deviations from the mean have been removed as outliers.

Source: Authors' calculations. • Created with Datawrapper

full-sample coefficients are marked with an asterisk.[10]

Figures 1-4 show the difference between standard TFP and the revised TFP based on our estimates. When the curve is below zero, the difference is negative, and TFP was underestimated in that period. Conversely, when the curve is above zero, TFP was overestimated. We begin by presenting the revised TFP curves for each individual intangible asset. To a certain extent, our results align with those of BRS (2021), even though we employ different data and methodologies, particularly to measure software and organizational capital.

estimations are similar to Figure 1 to 4.

Table 3: Estimated Coefficients Used for the Measurement of Productivity J-Curves

	France	Germany	Japan	United Kingdom	United States
The first type of J-curve (Figures 1 to 3)					
Research and development	8.53	2.84	2.32	3.56*	6.19
Software	1.28*	5.99	1.23	4.03	7.87
Organizational capital	1.96	4.79	3.08	3.80*	5.74
The second type of J-curve (combined intangibles case, Figure 4)					
Research and development	8.53	2.84	2.32	3.56*	6.19
Software	1.28*	5.99	1.23	4.03	7.87

Note: The table summarizes the coefficients used to compute the TFP measure revised accounting for all three intangible capital types at the same time, shown in Figure 4. Coefficients for R&D and software are obtained from the GMM estimates in Table 1, while those for organizational capital are obtained from the GMM estimates in Table 2. If any coefficient is not positive, significant, or larger than the respective asset price in those tables, it is replaced with the corresponding full sample estimate. Replaced coefficients are marked with an asterisk (*).

Source: Authors' calculations. • Created with Datawrapper

5.1 Research and Development

Figure 1 focuses on the effect of R&D. Our results for the U.S. are similar to BRS (2021) and in line with the arguments forwarded by Aghion *et al.* (2019). The R&D-adjusted TFP measure for the U.S. diverges only slightly from standard TFP, with differences of at most 0.4 percentage points, despite a slight underestimation towards the end of the sample. This pattern is similar for the other countries studied here, where the difference between the two TFP measures is even smaller — confirming that "intangible-related challenges for productivity estimation coming from R&D are likely to be minimal at present" (BRS, 2021).

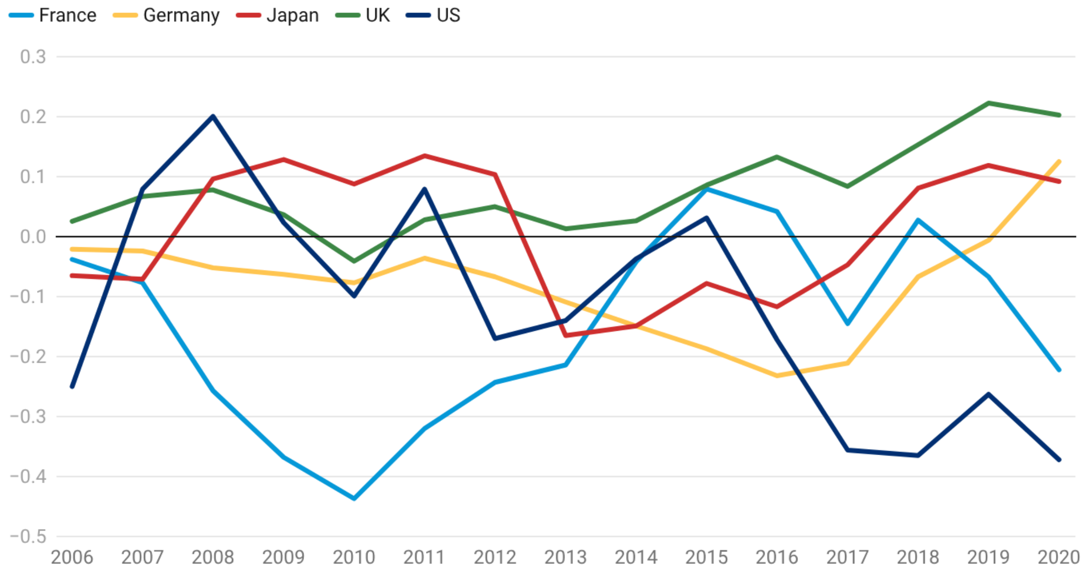
Specifically, we observe a slight underestimation of TFP growth in the U.S., particularly after 2010, and to an even lesser extent in France during the late 2000s and the early 2010s, and in Japan and Germany during the early 2010s. We do not find any underestimation in the United Kingdom.

5.2 Software

For software, a comparison with BRS (2021) is more challenging due to methodological differences. While they suggest a set of plausible values for the software-related coefficient, we estimate adjustment costs econometrically by creating a software investment variable based on both industry and firm level data. Figure 2 shows an under-estimation of TFP growth in the U.S. comparable in timing to BRS's

Figure 1: International Comparison of Productivity J-Curves for Research and Development

Difference between standard and revised TFP growth rates in percentage points



Note: The figure shows the 5-year moving average of the difference between standard and revised TFP growth. Revised TFP accounts for R&D-related intangible investment only. Values below zero indicate underestimation of TFP in standard measures.

Source: Authors' calculations. • Created with Datawrapper

finding, though less severe before 2015 and more since. However, our estimated magnitude of underestimation reaches 1.4 percentage points, which is significantly larger than the peak of 0.7 percentage points in the post-2006 period. This underestimation appears unique to the U.S., likely due to the large costs associated with software investment in the 2010s, which is consistent with the fact that many important digital innovations such as AI and platform businesses are developed in the United States.

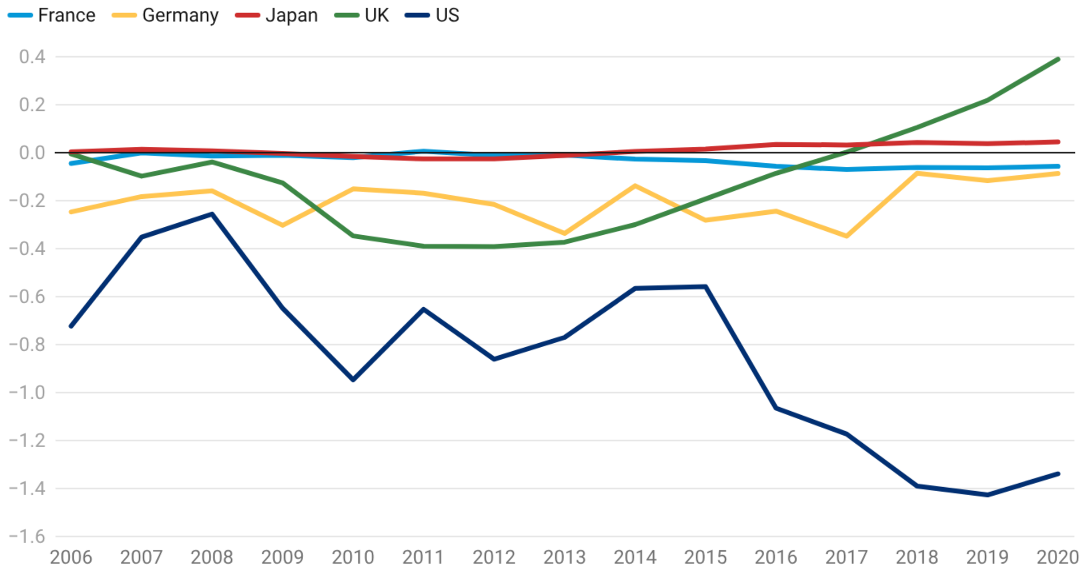
In Germany and in the UK, software-related TFP underestimation seems negligible, remaining below 0.4 percentage point for most of the period. It even turns positive in recent years in the UK, resembling a minor J-curve with smaller magnitudes, probably influenced by spillover effects from other countries. In Germany, the low undervaluation may reflect the lower intangible capital investment patterns of the country, especially in software (Nonnis, Roth and Bounfour, 2024), though a change is observable after 2018. Additionally, France and Japan show rather flat patterns, due to coefficients of less than 1 as obtained in Section 4, which do not allow the identification of unmeasured intangibles associated with software in these countries.

5.3 Organizational Capital

The results for organizational capital are more mixed, but suggest again the uniqueness of the U.S., where standard TFP growth is underestimated throughout most of the sample period, but without a J-curve type of pattern. However, the J-curve is

Figure 2: International Comparison of Productivity J-Curves for Software

Difference between standard and revised TFP growth rates in percentage points



Note: The figure shows the 5-year moving average of the difference between standard and revised TFP growth. Revised TFP accounts for software-related intangible investment only. Values below zero indicate underestimation of TFP in standard measures.

Source: Authors' calculations. • Created with Datawrapper

observable in the UK, where the difference between the two measures becomes positive after 2018, following a peak difference of 0.8 percentage points. In other countries, the revised TFP measure shows no substantial variations from standard TFP. However, the scale of mismeasurement by the associated costs in capital formation in organizational capital is relatively small compared to software. In particular, the underestimations of TFP growth rate caused by the associated costs of investment in organizational capital in continental Europe and Japan are smaller than those in the case of the UK and the United States.

These results imply that the firms in continental European countries such as France and Germany and Japan seem to be more reluctant to invest in organizational capital with large associated costs than those in the United States.[11]

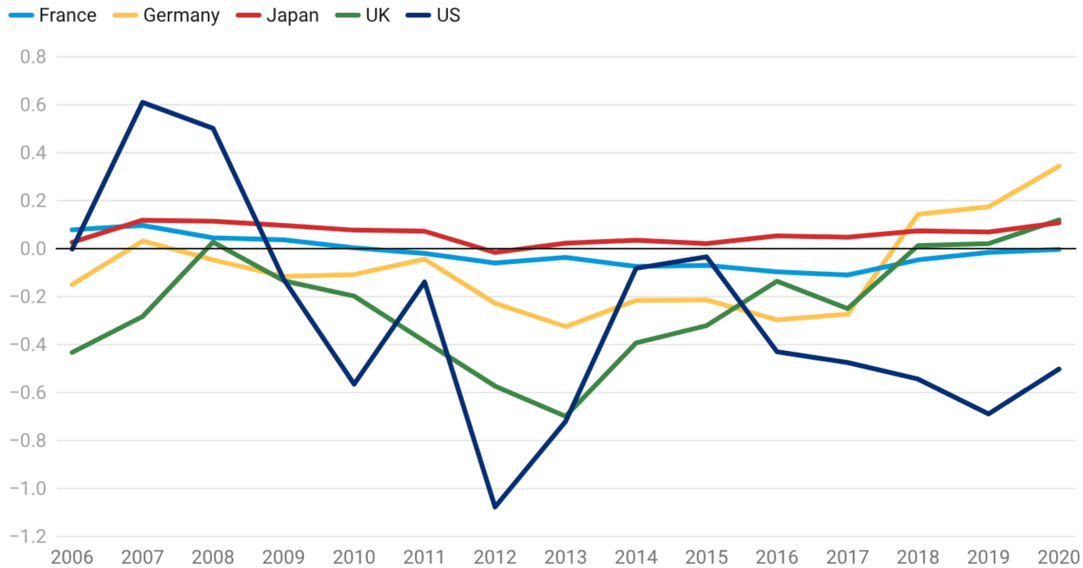
5.4 Combined Effects of All Intangibles

To produce a revised TFP measure that accounts for the additional intangible investment costs associated with R&D, software and organizational capital together, we use a combination of GMM coefficients from Tables 1 and 2, as summarized in Table 3. These coefficients are obtained from the GMM estimations in Table 1, except for organizational capital, where we

11 Although we conduct an alternative estimation using the measure of organizational capital developed by Eisfeldt and Papanikolaou (2013), our main conclusions do not change.

Figure 3: International Comparison of Productivity J-Curves for Organizational Capital

Difference between standard and revised TFP growth rates in percentage points



Note: The figure shows the 5-year moving average of the difference between standard and revised TFP growth. Revised TFP accounts for organizational capital-related intangible investment only. Values below zero indicate underestimation of TFP in standard measures.

Source: Authors' calculations. • Created with Datawrapper

use country-level GMM estimates from Table 2. As before, if any coefficient is not positive, significant, or larger than the respective asset price, we replace it with the corresponding full sample estimate instead. These replaced coefficients are marked with an asterisk (*) in Table 3.

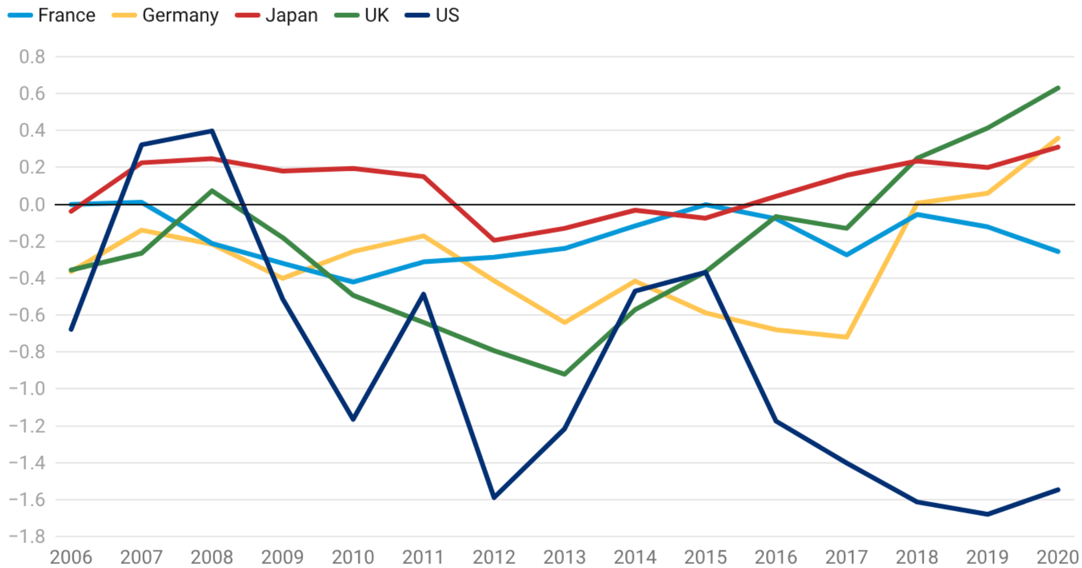
Figure 4 presents the combined effects of intangible correlates associated with three intangibles. Comparing these effects offers better insights into the impact of GPTs like AI on TFP, as it more accurately reflects the next generation of investments, because AI investment incorporates elements of all three intangibles we considered (Fonteneau *et al.*, 2025).[12]

Figure 4 again highlights the uniqueness of the U.S., as the under-estimation of TFP is present for almost the entire period, even reaching 1.6 percentage points in 2019. This suggests that the U.S. remains in a phase of rapid intangible investment growth, with positive effects on standard measures of TFP to be more than offset by further new investments. Germany, France and the UK seem to follow similarly, but with much smaller magnitudes. This implies the presence of some J-curve effects in these countries, though their impact is much smaller. Consistent with Miyagawa, Tonogi and Ishikawa (2021), the Japanese curve seems to be flat except for the minor underestimations in the early 2010s.

12 Although they argue the share of skills related to AI is the largest in the measurement of AI, we are not able to capture this component from the firm-level data.

Figure 4: International Comparison of Productivity J-Curves for All Intangibles

Difference between standard and revised TFP growth rates in percentage points



Note: The figure shows the 5-year moving average of the difference between standard and revised TFP growth. Revised TFP accounts for R&D, software and organizational capital-related intangible investments. Values below zero indicate underestimation of TFP in standard measures.

Source: Authors' calculations. • Created with Datawrapper

6. Conclusions

Productivity growth has stagnated in advanced economies in recent decades, particularly the 2010s. Mismeasurement of intangible capital has been identified as a potential explanation for this productivity slowdown. Productivity may have been underestimated because the additional value added and higher productivity induced by new technologies, such as AI and platform-based businesses, are not fully captured by current statistics. This is partly due to what BRS (2021) call *intangible correlates*, which are intangible assets associated with new technologies that are difficult to evaluate and account for.

Following BRS (2021), who constructed TFP measures corrected for the adjustment costs of intangibles, and observed that the gap between standard and revised TFP formed a J-shaped pattern (the productivity J-curve), this paper extends their analysis to European countries and Japan, which also experienced a productivity slowdown despite digitalization. Using firm-level data from Orbis, covering listed firms, and industry-level data from sources such as EUKLEMS/INTANProd and the Japanese Industrial Productivity (JIP) database, we construct revised TFP measures and assess whether the gap between standard and revised measures followed a productivity J-curve pattern not only in the U.S., but also in France, Germany, the UK and Japan.

Our revised measures are adjusted by three types of intangible capital (R&D, software and organizational capital), considered both individually and aggregately.

In line with what was observed by BRS (2021) for the U.S., we find that while the effects are small and negligible for R&D, they are clear for software and organizational capital, but mostly limited to the United States.

In particular, we show that most of the observed effects come from software in the United States. This result is consistent with the fact that many important digital innovations, such as AI and platform businesses, are developed in the United States. As the costs associated with software investment are not reflected as intangibles in the standard productivity statistics, the standard measure of TFP growth is underestimated by up to 1.4 percentage points, and 1.6 percentage points when considering all three intangibles. This result is in line with the recent productivity trends observed in the United States. The Bureau of Labor Statistics Report (BLS, 2026) issued in March 2026 shows an annualized growth rate of 2.1 per cent in the business cycle starting in the fourth quarter of 2019, higher than that observed in the previous business cycle (1.5 per cent). We argue that this acceleration of productivity growth in the U.S. has been driven by digitalization in the service sector, which is compatible with the underestimations observed in our measures.

We find the underestimation of TFP growth rate caused by the associated costs of investment in organizational capital to be small in the continental European countries and Japan. These results imply the management practices in continental Europe and Japan seem to have caused firms to be more reluctant to invest in organizational capital than in the United States.

Our study shows that the effect of unaccounted intangibles on the productivity slowdown in advanced countries during the 2010s differs by country. We do not find a large underestimation of TFP growth associated with capital formation in intangibles in the continental European countries and Japan, while there is still large underestimation of TFP growth rate in the U.S. in the late 2010s. This implies the productivity gap after taking account of the adjustment costs of investment between the U.S. and other advanced countries is even larger than measured by the standard TFP statistics.

These results do not directly reflect differences in investment levels among the countries in our sample. Countries that appear to invest heavily in national accounts data are not necessarily those that have high intangible correlates. On the contrary, countries that invest less in official statistics may have more intangible correlates if the lower investment is the result of the measurement errors or cross-country methodological disharmonizations highlighted in the cited papers: national accounts do not record those intangibles, but the BRS method captures them through market value regressions. In other words, financial markets value "invisible" assets even when national accounts fail to record them. This implies that unmeasured intangible correlates of different magnitudes in each country are captured by the adjustment costs in our econometric estimations. This discrepancy between investment data and intangible intensity reflects measurement difficulties that are evident when examining official national statistics and EUKLEMS/INTANProd data.

For example, according to these data, Germany has invested considerably less (10.7 per cent of value added) compared to other European countries and the U.S. in recent decades, while France has invested much more (16 per cent of value added), reaching levels comparable to the U.S. (16.3 per cent). These stylized facts reflect potential measurement issues pointed out in Nonnis, Roth and Bounfour (2024, 2025), which the BRS methodology overcomes.

The absence of a clear J-curve pattern in some high investing countries should be considered further. Nevertheless, it reinforces the call for a more focused investment, particularly in new technologies, as emphasized by the recent European Commission report by Mario Draghi (2024a, 2024b), which highlights that Europe is lagging behind the U.S. and China in this respect. Our results suggest that the productivity gap arising from those differences in investment behavior may be even larger, as official productivity growth figures for the U.S. may have been underestimated due to intangible capital mismeasurements, an effect we do not observe as strongly in European countries and Japan.

While it is difficult to predict exact patterns for the next generation of investments, including AI, some lessons can be learned by looking at recent AI investment trends across countries. According to the Artificial Intelligence Index report (Maslej *et al.*, 2024), the U.S. leads in private AI investment, investing nearly twice as much as in the UK relative to its GDP over the last ten years and at least four times more than in the other countries. This suggests that J-curve effects due to AI will likely be much stronger in the U.S. than in other countries, as a result of both the uniqueness of the U.S. highlighted in this study, and its superior investment levels in AI. The intensive investment in AI in the U.S. promises a higher TFP growth rate than other advanced countries as Filippucci, Gal and Schief (2024) estimated. To catch up, continental European countries and Japan should invest more aggressively in digital innovation and in the associated costs of innovative capital formation, such as training skilled workers (Bounfour, Nonnis, Yang, 2025), while also reforming conservative management practices.

References

Acemoglu, D., D. Autor, D. Dorn, G.H. Hanson, and B. Price (2014) "Return of the Solow Paradox? IT, Productivity, and Employment in U.S. Manufacturing," *American Economic Review*, Vol. 104, No. 5, pp. 394–399.

Ackerberg, D.A., K. Caves, and G. Frazer (2015) "Identification Properties of Recent Production Function Estimators," *Econometrica*, Vol. 83, No. 6, pp. 2411–2451.

Aghion, P., A. Bergeaud, T. Boppart, P. Klenow, and H. Li (2019) "Missing Growth from Creative Destruction," *American Economic Review*, Vol. 109, No. 8, pp. 2795–2822.

Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, Vol. 58, No. 2, pp. 277–297.

Arellano, M. and O. Bover (1995) "Another Look at the Instrumental-Variable Estimation of Error-Components Models," *Journal of Econometrics*, Vol. 68, No. 1, pp. 29–51.

Basu, S., J.G. Fernald, N. Oulton, and S. Srinivasan (2003) "The Case of the Missing Productivity Growth," in M. Gertler and K. Rogoff (eds.), *NBER Macroeconomics Annual*, MIT Press, Cambridge, MA, pp. 9–82.

Bavdaž, M., A. Bounfour, J. Martin, A. Nonnis, G. Perani, and T. Redek (2023) "Measuring Investment in Intangible Assets," in G. Snijkers, M. Bavdaž, S. Bender, J. Jones, S. MacFeely, J.W. Sakshaug, K.J. Thompson, and A.v. Delden (eds.), *Advances in Business Statistics, Methods and Data Collection*, Wiley. Available at: https://doi.org/10.1002/9781119672333.ch5.

Bijnens, G., J. Konings, and A. Putseys (2025) "Unveiling the J-Curve: How Intangibles Drive Productivity Mismeasurement," CEPR Discussion Paper, No. 20825.

BLS (U.S. Bureau of Labor Statistics) (2026) *Productivity and Costs. Fourth Quarter and Annual Averages 2025, Revised*, U.S. Department of Labor. Available at: https://www.bls.gov/news.release/prod2.nr0.htm.

Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 87, No. 1, pp. 115–143.

Bontadini, F., C. Corrado, J. Haskel, M. Iommi, and C. Jona-Lasinio (2023) "EUKLEMS and INTANProd: Industry Productivity Accounts with Intangibles," MPRA Paper, No. 126218, University Library of Munich.

Bontadini, F., C. Corrado, J. Haskel, M. Iommi, C.S. Jona-Lasinio, and T. Miyagawa (2024) "Intangible Capital, TFP Growth and Green Shoots in New Productivity Data," *International Productivity Monitor*, No. 46, pp. 3–37.

Bounfour, A. and A. Nonnis (2025) "Business Agility, Intangible Capital, and Economic Performance during the Covid-19 Crisis," *Science and Public Policy*, Vol. 52, No. 6, pp. 941–949.

Bounfour, A., A. Nonnis, and S. Yang (2025) "Assessing the Impact of AI Skills on Firm Productivity: A Cross-country Analysis," *Digital Transformation and Society*, Vol. 4, No. 3, pp. 299–311.

Brynjolfsson, E., L.M. Hitt, and S. Yang (2002) "Intangible Assets: Computers and Organizational Capital," *Brookings Papers on Economic Activity*, No. 1, pp. 137–198.

Brynjolfsson, E., D. Rock, and C. Syverson (2019) "Artificial Intelligence and the Modern Productivity Paradox," in A. Agrawal, J. Gans, and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago, pp. 23–60.

Brynjolfsson, E., D. Rock, and C. Syverson (2021) "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, Vol. 13, No. 1, pp. 333–372.

Corrado, C., C. Hulten, and D. Sichel (2009) "Intangible Capital and U.S. Economic Growth," *Review of Income and Wealth*, Vol. 55, No. 3, pp. 661–685.

Coyle, D. and L. Nakamura (2022) "Time Use, Productivity, and Household-Centric Measurement of Welfare in the Digital Economy," *International Productivity Monitor*, No. 42, pp. 165–186.

Draghi, M. (2024a) *The Future of European Competitiveness: Part A*, European Commission.

Draghi, M. (2024b) *The Future of European Competitiveness: Part B*, European Commission.

Eisfeldt, A. and D. Papanikolaou (2013) "Organizational Capital and the Cross-Section of Expected Returns," *Journal of Finance*, Vol. 68, No. 4, pp. 1365–1406.

Fernald, J., R. Inklaar, and D. Ruzic (2025) "The Productivity Slowdown in Advanced Economies: Common Shocks or Common Trends?" *Review of Income and Wealth*, Vol. 71, No. 1, Article e12690.

Filippucci, F., P. Gal, and M. Schief (2024) "Miracle or Myth? Assessing Macroeconomic Gains from Artificial Intelligence," OECD Artificial Intelligence Papers, No. 29.

Fonteneau, F., J. Mollins, S. Marchi, L. Russo, A. Gentaz, M. Daoud, and A. Andrè (2025) "Advancing the Measurement of Investments in Artificial Intelligence," OECD Artificial Intelligence Papers, No. 47.

Goldin, I., P. Koutroumpis, F. Lafond, and J. Winkler (2024) "Why Is Productivity Slowing Down?" *Journal of Economic Literature*, Vol. 62, No. 1, pp. 196–268.

Gordon, R.J. (2016) *The Rise and Fall of American Growth*, Princeton University Press, Princeton, NJ.

Hall, R. (2001) "The Stock Market and Capital Accumulation," *American Economic Review*, Vol. 91, No. 5, pp. 1185–1202.

Hasegawa, S. (2023) "How Can the Digital Economy Be Grasped?" *The Economic Analysis*, No. 207, pp. 250–276 (in Japanese).

Hayashi, F. (1982) "Tobin's Marginal q and Average q," *Econometrica*, Vol. 50, No. 1, pp. 213–224.

Hayat, Z. (2022) "Digital Trust: How to Unleash the Trillion-Dollar Opportunity for Our Global Economy," *World Economic Forum*.

Hulten, C. and X. Hao (2008) "What Is a Company Really Worth?" NBER Working Paper, No. 14548. Available at: https://www.nber.org/papers/w14548.

Lev, B. and S. Radhakrishnan (2003) "The Measurement of Firm-Specific Organization Capital," NBER Working Paper, No. 9581. Available at: https://www.nber.org/papers/w9581.

Lucas, R.E., Jr. (1967) "Adjustment Costs and the Theory of Supply," *Journal of Political Economy*, Vol. 75, No. 4, pp. 321–334.

Maslej, N., L. Fattorini, R. Perrault, V. Parli, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J.C. Niebles, Y. Shoham, R. Wald, and J. Clark (2024) "The AI Index 2024 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. Available at: https://hai.stanford.edu/ai-index/2024-ai-index-report.

Miyagawa, T. (2024) "Economics of Digitalization," *The Economic Analysis*, No. 209, pp. 1–25 (in Japanese).

Miyagawa, T. and Y.G. Kim (2008) “Measuring Organization Capital in Japan,” *Seoul Journal of Economics*, Vol. 21, No. 1, pp. 169–193.

Miyagawa, T., M. Takizawa, and K. Edamura (2015) “Does the Stock Market Evaluate Intangible Assets?” in A. Bounfour and T. Miyagawa (eds.), *Intangibles, Market Failure, and Innovation Performance*, Springer, Heidelberg, pp. 113–138.

Miyagawa, T., K. Tonogi, and T. Ishikawa (2021) “Does the Productivity J-Curve Exist in Japan?” *Journal of the Japanese and International Economies*, Vol. 61, Article 101126.

Nonnis, A., A. Bounfour, and K. Kim (2023) “Knowledge Spillovers and Intangible Complementarities: Empirical Case of European Countries,” *Research Policy*, Vol. 52, No. 1.

Nonnis, A., F. Roth, and A. Bounfour (2024) *Is the EU Ready for the Next Generation of Investment?*, Hamburg Discussion Papers in International Economics, No. 16.

Nonnis, A., F. Roth, and A. Bounfour (2025) “Intangible Capital in France and Germany: Is There a Measurement Issue?” *AStA Wirtschafts- und Sozialstatistisches Archiv*. Available at: https://doi.org/10.1007/s11943-025-00358-4.

Tambe, P., L. Hitt, D. Rock, and E. Brynjolfsson (2020) *Digital Capital and Superstar Firms*, NBER Working Paper, No. 28285. Available at: https://www.nber.org/papers/w28285.

Uzawa, H. (1969) “Time Preference and the Penrose Effect in a Two-Class Model of Economic Growth,” *Journal of Political Economy*, Vol. 77, No. 4, pp. 628–652.

van Ark, B., K. de Vries, and A. Erumban (2024) “Are Intangibles Running out of Steam?” *International Productivity Monitor*, No. 46, pp. 38–59.

Wildasin, D. (1984) “The q Theory of Investment with Many Capital Goods,” *American Economic Review*, Vol. 74, No. 1, pp. 203–210.

Does the Import Invasion Explain the Disappearance of Productivity Growth in U.S. Manufacturing?

Robert J. Gordon and Kenneth Ryu*

IPM Research Article

Abstract

Why did U.S. manufacturing productivity stop growing after 2010? Productivity growth actually disappeared, from an annual rate of +3.3 per cent during 1987-2010 to -0.3 per cent from 2010 to 2023. This article shifts attention from 2010 as the start of the productivity growth slowdown to a decade earlier when *output* stopped growing. This cessation of output growth in 2000 is attributed to the invasion of imports that closed domestic plants, destroyed jobs, and squeezed profits. After 2000, a chain of causation followed that ultimately undermined productivity growth — from falling capacity utilization, to lower investment in fixed capital and research and development, and an erosion of innovation. Beyond the import invasion, the disappearance of productivity growth is also attributed to a general phenomenon of diminishing returns to innovation, the feeble influence of robots, government regulations that distorted investment, and a shrinking supply of skilled labour in the face of increasing skill demands.

1. Introduction

While the manufacturing sector produces only 10 per cent of U.S. gross domestic product (GDP), it accounts for more than half of U.S. research and development (R&D) and produces most of U.S. exports, fixed investment, and consumer goods. Yet the former dynamism of that sector appears to have vanished. The growth rate of labour productivity in U.S. manufacturing in official data — produced by the Bureau of Labor Statistics (BLS) and Bureau of Economic Analysis (BEA) — was 3.3 per cent per year over the 23 years between 1987 and 2010. But then growth disappeared, registering a negative rate of -0.3 per cent over the following 13 years

* Robert J. Gordon, Northwestern University. Email: rjg@northwestern.edu. Kenneth Ryu, Northwestern University. Email: KennethRyu@u.northwestern.edu. We are grateful to Bart van Ark for helpful comments, to David Byrne for suggestions on price deflation for the computer industry, to Chad Syverson for sharing the detailed industry data on the Atalay *et al.* price adjustments, and to John Ma for superb background research.

1 Labour productivity is gross output from the BEA National Income and Product Accounts “GDP by Industry” tables and hours are from the BLS productivity table “Total Factor Productivity by Major Industries.” The

between 2010 and 2023.[1] The vanishing of productivity growth in U.S. manufacturing raises deep questions about the historical core of American prosperity.

How could productivity growth in the entire manufacturing sector completely disappear? Has U.S. manufacturing lost its ability to compete? A substantial literature has emerged that examines numerous possible causes of the productivity slowdown. But those papers contrast productivity growth before and after 2010, with little or no attention to profound changes that occurred before 2010. Yet the deep malaise of U.S. manufacturing began at least a decade earlier. While productivity growth was zero after 2010, manufacturing gross output in the year 2000 began a 23-year interval of zero annual growth.[2] That is, the level of manufacturing gross output (which includes the value of intermediate inputs) was the same in 2023 as in 2000, while the level of real GDP (which nets out the value of intermediate inputs) was 57 per cent higher, a stunning contrast.

By what channels does a flood of imported goods undermine the efficiency and viability of domestic manufacturing firms? We find that there is a strong negative correlation across the 19 three-digit manufacturing industries between the post-1987 rise in import penetration and the subsequent post-2000 stagnation of domestic output. Foreign competition not only causes plant closures in the most affected industries like furniture, textiles, and apparel, but squeezes profits throughout other parts of the manufacturing sector, thus cutting funding for fixed investment as well as R&D. There is a striking shift from positive to negative growth of investment after 2000 in many manufacturing industries. Further, when innovation originates from foreign producers of imports, domestic firms lose touch with leading-edge technology, are less able to achieve their own technological advances, and are less able to obtain critical inputs from domestic sources.

There are limits to the import/offshoring explanation of the productivity growth slowdown. By 2018, import penetration (defined as imports divided by imports plus domestic production) had reached 30 per cent or above in half of the manufacturing industries. Averaged across the 19 industries, import penetration increased from 13 per cent in 1988 to 28 per cent in 2018; in the latter year import penetration ranged all the way from 5 per cent in petroleum to 97 per cent in apparel. Yet the productivity growth slowdown was larger for petroleum than for apparel, suggesting that offshoring and imports provide only a partial explanation of productivity growth patterns.

We broaden the investigation by examining symptoms of a deeper malaise that extends beyond manufacturing into the rest of the American economy. A recent New York Times editorial reports decades during which the American shipbuilding industry has delivered ships billions of dollars over budget, years behind schedule, that fail to perform as specified. The U.S.

output data are subsequently adjusted below for alternative price deflators.

2 The 2000-23 annual growth rate of manufacturing output in the BEA data is -0.05 per cent.

aircraft industry takes 12 years to deliver a new military fighter or bomber. "America's defense industry, like much of the economy, has lost the ability to build quickly and effectively" (New York Times Editorial Board, 2025).

This article begins with measurement issues. A recent paper by Atalay *et al.* (2025) convincingly argues that the Producer Price Index (PPI) used by the BEA to deflate output in a given category of goods systematically rises faster than the corresponding Consumer Price Index (CPI), due at least in part to more extensive correction for quality change in the CPI. We incorporate their price adjustments for all of the 19 industries except the computer and electronics industry, where we rely instead on the original research of David Byrne and his coauthors. Because the Atalay-Byrne price adjustments are of roughly the same magnitude before and after 2010, they do not explain any of the productivity growth slowdown, in fact the reverse.

We differ from the previous literature along two other dimensions. Unlike previous papers, including Atalay *et al.* (2025), that examine the slowdown of total factor productivity (TFP) growth, we focus instead on labour productivity growth, defined as growth in gross output per hour instead of real value-added per hour. Several previous papers have shown that the difference between these two, i.e., intermediate inputs, is subject to "offshoring bias" that can cause the growth of inputs to be understated, and hence the derived growth of value added and TFP to be overstated (Houseman *et al.* 2011). A second difference with other research is that we choose 2005 rather than 2010 as the break year to define the slowdown for reasons related to the timing effects of the 2008-09 recession.

The article begins with measurement issues and continues with an examination of differences in the magnitude of the slowdown across the 19 manufacturing industries. We then study the channels by which growing import penetration contributed to the cessation of output growth after 2000 and productivity growth after 2010. Next, we examine the sources of declining innovation as a side-effect of shrinking public R&D, changing corporate strategies about private R&D, and the evidence of diminishing returns to R&D in creating innovation. Finally, we examine three additional causes of the slowdown, including the robot puzzle; the role of government regulations in distorting investment and reducing productivity; and the persistent shortage of skilled labour in manufacturing.

2. Measurement: Inflation Bias, Break Date, and Offshoring Bias

2.1 Price Deflator Bias

For many decades economists have demonstrated that price indexes — particularly for durable goods — overstate inflation by failing to take account of quality change and the value of new products. A comprehensive set of new measures of durable goods prices was provided by Gordon (1990). Gradually over the years the BLS has incorporated into its price indexes better measures of quality change, but evidence is not usually available to extend the quality adjustments backwards in time.

This leads to a presumption that the extent of upward inflation bias in deflators, and the resulting downward bias in the growth of real output, diminishes over time. Ironically this measurement issue would tend to make pre-2010 productivity growth rates too low, since less accurate price measurement would make pre-2010 output growth rates understated more than their post-2010 counterparts.

A recent attempt partially to address the inflation bias issue has been provided in a much-noticed paper by Atalay *et al.* (2025). They argue that the BLS devotes more resources to making quality adjustments in the CPI than the PPI and Import Price Index that are currently used to deflate gross output and intermediate inputs. To reduce this quality change bias, the authors carry out the complex task of using input-output tables to match CPI products with specific industries. We accept the Atalay corrections for all of the 19 industries except for the computers and electronic products industry.[3] For that we adopt price change corrections suggested by David Byrne based on research by himself and coauthors.[4] The Byrne corrections for computers and electronics alter the growth rate of the BEA deflators by -6.9 per cent per year for 1987-2010 and by -5.4 per cent for 2010-23.

For total manufacturing, the CPI substitution raises output growth by 1.6 per cent per year for 1987-2010 and by 1.0 per cent for 2010-23. The adjustments for durable goods are 2.7 and 2.2 per cent respectively, much of which is due to the computer industry. When this industry is excluded the durable goods adjustments drop to 1.8 for both intervals. The nondurable goods adjustments are 1.0 and 0.5 per cent. Except for durables excluding computers, all of these adjustments are smaller for post-2010 than pre-2010, implying that the Atalay revisions do not help explain the productivity growth slowdown at all, but rather make it slightly larger. What they do accomplish is to eliminate, if only slightly, the puzzle of "disappearing" productivity growth in manufacturing, as the 2010-23 productivity growth rate is boosted from -0.3 to +0.7 per cent per year. All productivity growth numbers cited in this article henceforth use only the Atalay-Byrne adjusted data.[5]

2.2 Implications of Offshoring Bias

As emphasized in the 1996 Boskin Commission report, a source of upward bias in the CPI has long been "outlet substitution bias." When the price quote for a given product sold at a full-price merchant like Macy's is replaced by the lower price sold at a rising discount merchant like Walmart, the CPI methodology links out the price decrease and so the benefit to the consumer

3 Chad Syverson kindly provided detailed annual price change adjustments by industry for 1998-2023. We extended these by applying the average adjustment from 1998-2005 uniformly to each year between 1988 and 1997.

4 Byrne's papers provide suggested corrections to BEA deflators for industry 334 for three time intervals: 1978-95, 1995-2004, and 2004-14. We have extended his suggested bias corrections to 2023; details of this translation are provided in the Data Appendix.

5 Table 1 that follows compares the two alternative series.

is not captured in the index. Houseman *et al.* (2011) identified a parallel price index bias that occurs for the same reason but, rather than involving domestic full-price and discount merchants, instead contrasts intermediate materials previously sold by high-cost domestic producers which are replaced by lower-cost imported materials. The price quotes for the former high-cost domestic producer are not directly linked to the new lower price quote from the foreign supplier. Thus the actual price decline is missed in the deflator for intermediate materials.

This "offshoring bias" leads to an upward bias in the price change of intermediate goods and corresponding downward bias in the growth of real intermediate materials inputs. This matters because real gross output change is based on original data (subject to the price measurement issues discussed above). But everything else in growth accounting is a derived number. If offshoring bias is significant, then the growth of real value added (RVA, that is, gross output growth minus intermediate materials growth) is overstated. Because TFP growth is calculated as RVA growth minus the contributions of capital deepening and labour composition, TFP growth is overstated as well. To the extent that capital input production shifts from domestic to foreign sources, the upward bias in TFP measures is compounded.

Most of the recent literature on the productivity slowdown, including the Atalay paper, focuses on comparisons of TFP growth before and after 2010, and thus is subject to this source of overstatement of TFP growth. Whether this makes the TFP slowdown greater or smaller depends on whether offshoring bias was more or less important after 2010. One approach to assess its impact is to compare the growth of RVA which is altered by offshoring bias and gross output which is not. In our Atalay/Byrne-adjusted data RVA grows faster than gross output by 1.2 per cent per year in 1987-2010 and 0.8 per cent faster during 2010-23, so offshoring bias explains 0.4 annual percentage points of the slowdown in TFP growth leaving aside anything extra from capital deepening. To avoid the inaccuracy introduced by offshoring bias, our article examines only changes in labour productivity defined as gross output per hour and makes no further mention of RVA or TFP growth.

2.3 Break Year

We choose a different break year for defining the slowdown. Most of the recent literature uses 2010 as the break year, comparing 1987-2010 or 1997-2010 with 2010-23. This is understandable, since the puzzle of near-zero productivity growth begins when average growth rates are calculated starting with the 2010-11 annual change. That choice skews the outcome of the slowdown investigation, because the literature generally ignores the fact that the annual rate of change of labour productivity for the single year change of 2009-10 was 8.0 per cent for total manufacturing and 11.6 per cent for durable goods.

Gordon-Sayed (2025) have examined this jump in productivity in quarterly data where the large positive growth occurs in the last three quarters of 2009. Their explanation is that the rapid collapse of output in the fall of 2008, particularly in durable goods manufacturing, led firms to overreact and reduce work hours with a lag by a greater percentage than the decline in output, differing from previous recessions when the percentage change in hours was always less than in output. In their study of the business sector, they show that the "extra layoffs" of 2008-09 and corresponding temporary jump in productivity growth were followed by a rehiring reversal starting in 2010. They conclude that this artifact of the 2008-09 recession leads productivity growth to be overstated in 2008-09 in quarterly data and correspondingly understated in 2010-19 compared to the underlying determinants of trend productivity growth.

To avoid this cyclical distortion, we define the slowdown with 2005 as the break year instead of 2010. Given our desire to avoid any effect of the 2008-09 recession, we could have chosen 2007 instead of 2005. However, 2005 has the appeal that it makes our pre/post slowdown intervals an equal 18 years in duration (1987-2005 compared with 2005-23). The selection of 2005 as the cutoff year is reinforced by the previous research that identifies a structural break in U.S. TFP growth around 2004–2005 (2004Q4), marking the end of the late-1990s/early-2000s ICT-driven surge (Fernald, 2015).

The effects of our price adjustment and break year choices are summarized in Table 1. The top frame uses the break year 2010 for the official BEA/BLS growth rates. Two lines are shown for total manufacturing with and without the computer and electronic products industry. All productivity growth rates are defined as the average annual growth rate of gross output for a given interval minus the corresponding growth rate of labour hours.

In the top frame the two growth slowdowns are 3.6 and 2.9 per cent per year, respectively. We will find throughout the article that when the computer industry is excluded the magnitude of the productivity growth slowdown is substantially reduced. The 2010-23 productivity growth rates are negative for both lines in the top frame.

The next frame of Table 1 shows the same growth rates after the Atalay/Byrne price deflator adjustments are applied. The pre/post 2010 growth rates are boosted by about the same amount and so the magnitude of the slowdown is changed only slightly by +0.5 and +0.2 percentage points respectively. That is, the alternative price deflators *raise* the magnitude of the slowdown for both rows. The post-2010 growth rates of productivity are boosted enough to become positive rates. Thus, the price adjustments "solve" the puzzle of why productivity growth in manufacturing completely disappeared after 2010 but do not shed any light on the magnitude of the slowdown.

The bottom frame of Table 1 calculates the same growth rates when the break year is switched from 2010 to 2005. The growth rates for 1987-2005 are slightly higher than for 1987-2010, simply because the growth rates in the 2005-10 subinterval are lower than the pre-2005 growth rates. But the 2005-10 growth rates are substantially higher than for 2010-23, raising the resulting 2005-23 growth rates by between 0.9 and 0.7 per cent annually compared to 2010-23 for total manufacturing with and without the computer/electronic industry, respectively.

Table 1: Annual Average Growth Rates of Labour Productivity, by Different Price Deflators, Break Dates and Industry Components

Per cent

Panel A	1987-2010	2010-2023	Slowdown*
BLS/BEA data			
Total manufacturing	3.3	-0.3	-3.6
Total manufacturing excluding computers	2.4	-0.5	-2.9
Panel B	**1987-2010**	**2010-2023**	**Slowdown***
Price-adjusted data			
Total manufacturing	4.9	0.7	-4.1
Total manufacturing excluding computers	3.4	0.3	-3.1
Panel C	**1987-2005**	**2005-2023**	**Slowdown***
Price-adjusted data			
Total manufacturing	5.1	1.6	-3.5
Total manufacturing excluding computers	3.5	1.0	-2.5

Notes: Labour productivity growth rates are defined as the average annual growth rate of gross output minus the growth rate of labour hours over the same period. * denotes percentage points.

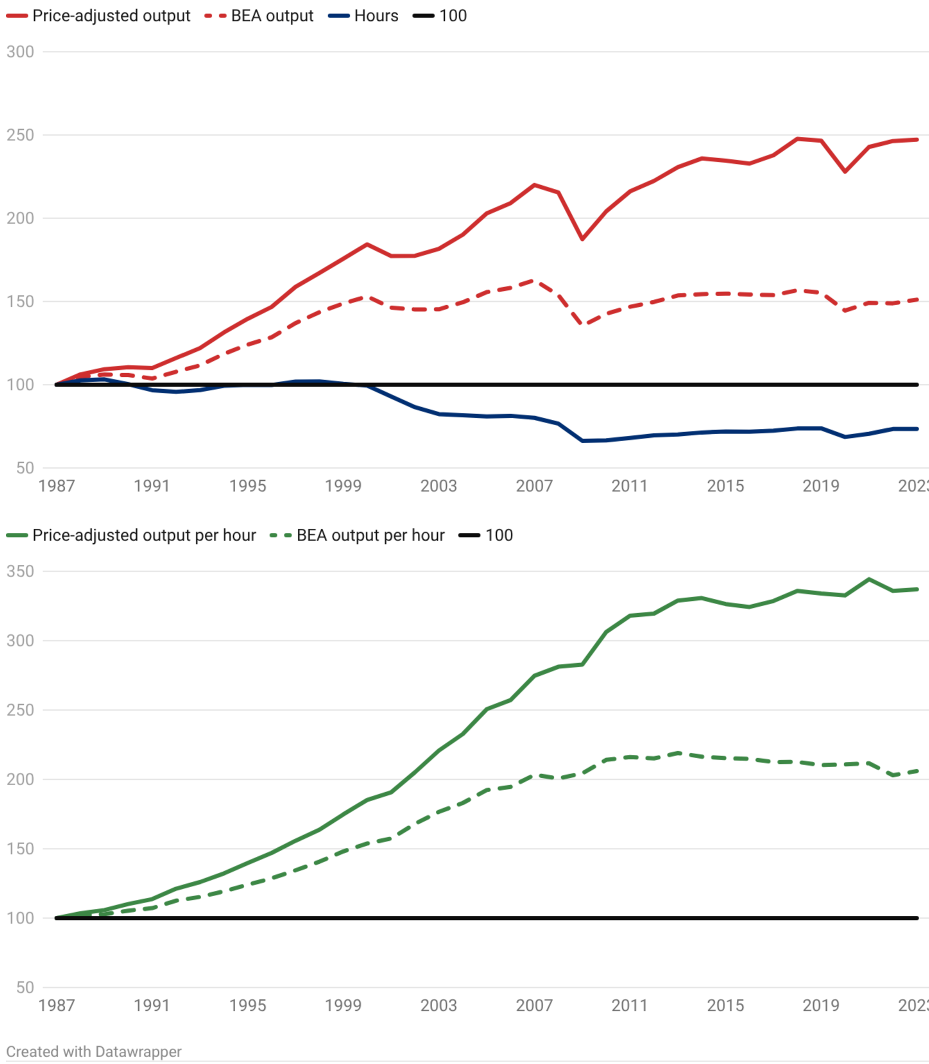
Source: Authors' calculations. • Created with Datawrapper

The switch of break dates accomplishes a small part of our goal to explain as much as possible of the productivity growth slowdown. The choice of the 2005 break date reduces the slowdown by 0.6 percentage points for both rows. Overall, the price-adjusted slowdown for total manufacturing of 4.1 percentage points is reduced to 3.5 points by switching the break year, and to 2.5 points by excluding the computer industry. These two changes achieve a 39 per cent reduction in the overall magnitude of the slowdown ((4.1-2.5)/4.1).

How is the evolution of labour productivity divided between its output numerator and hours denominator? The top frame of Figure 1 displays index numbers (1987=100) for output in red and hours in blue, with the corresponding index number for labour productivity as the green lines

Figure 1: Gross Output, Hours, and Output per Hour, Total Manufacturing, 1987-2023

Indexed to 100 in 1987



in the bottom frame. Dashed lines represent the official BEA/BLS series for output and productivity, while the solid lines plot the series after applying the Atalay/Byrne price adjustments. By depicting the year-by-year evolution of the data, Figure 1 provides additional insight into the timing of the phenomena that we need to explain.

The previous literature has focused on the post-2010 productivity growth slowdown but has generally not noticed another stark slowdown ten years earlier, that of output growth after 2000. The official BEA output series, shown by the dashed red line, stagnates after 2000, with a 2000-23 annual growth rate of -0.1 per cent, sharply down from the 3.3 per cent annual growth rate that occurred during 1987-2000. The price adjustments raise the 2000-23 growth rate of output slightly to a positive 1.3 per cent per year, but the sharp downturn from 1987-2000 to 2000-23 is roughly the same, from 4.7 to 1.3 per cent. We shall return to the theme below that the seeds for the near-disappearance of productivity change after 2010 were planted a decade earlier in 2000 when the import invasion began to swamp domestic industries.

Figure 1 also shows that the 2000-10 decade was different from either pre-2000 or post-2010. The price-adjusted output series shown by the solid red line grew at only 1.0 per cent per year, while employment collapsed, registering an annual growth rate of -4.0 per cent per year. Thus, the apparent similarity of productivity growth in 1987-2000 and 2000-2010, 4.7 and 5.0 per cent respectively, disguises a stark change from healthy output growth and stable employment before 2000 to stagnant output and rapidly collapsing employment during 2000-10. To the extent that the surge of imports caused the post-2000 transition, plant closings of low-productivity plants changed the mix within those industries to higher productivity plants and firms. This is part of the reason that productivity growth remained positive after 2000 through 2010. We return to this theme below.

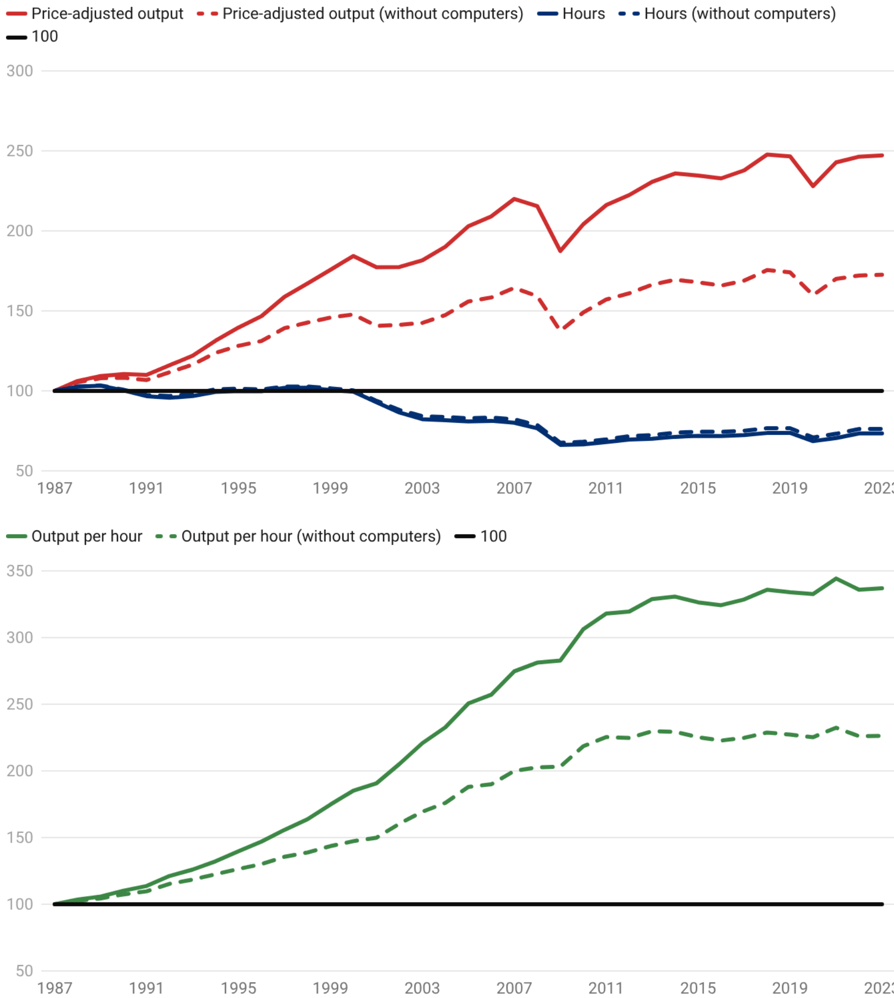
In Figure 2 the solid lines are the same as in Figure 1, while the dashed red and green lines plot the price-adjusted index numbers for output and productivity when the computer industry is excluded. The most important element of Figure 2 is the dashed red line showing that when computers are excluded, output stagnated after 2000, achieving an annual growth rate between 2000 and 2023 of only 0.7 per cent per year after growing at 3.0 per cent per year during the 13 years prior to 2000. For the 2000-10 decade this measure of output growth was 0.1 per cent per year. Thus for 18 of the 19 manufacturing industries the phenomenon of complete stagnation of output growth after 2000 remains valid even with the price adjustments and becomes one of the most interesting facts that needs to be explained. To what extent does the import invasion explain this output stagnation, and does the disappearance of output growth help explain, in turn, why productivity growth transitioned after 2010, when prices are adjusted and computers are excluded?

The lower frame in Figure 2 contrasts the index of manufacturing productivity with and without computers. With the 2010 break date excluding the computer industry explains 24 per cent of the productivity growth slowdown, while using the alternative 2005 break date, the computer contribution is an almost identical 29 per cent.[6] Using conventional BEA/BLS data and a 1997-2023 time span with a 2011 break year, a recent report by Chittoor *et al.* (2025), found a computer industry contribution of 16 per cent.

6 For total manufacturing productivity growth slows from 5.1 to 1.6 per cent, for a slowdown of 3.5 per cent per year. With computers excluded those three numbers are 3.5, 1.0, and 2.5. The computer contribution is ((3.5-2.5)/3.5) or 29 per cent.

Figure 2: Gross Output, Hours and Output per Hour, Total Manufacturing with and without Computers, 1987-2023

Indexed to 100 in 1987



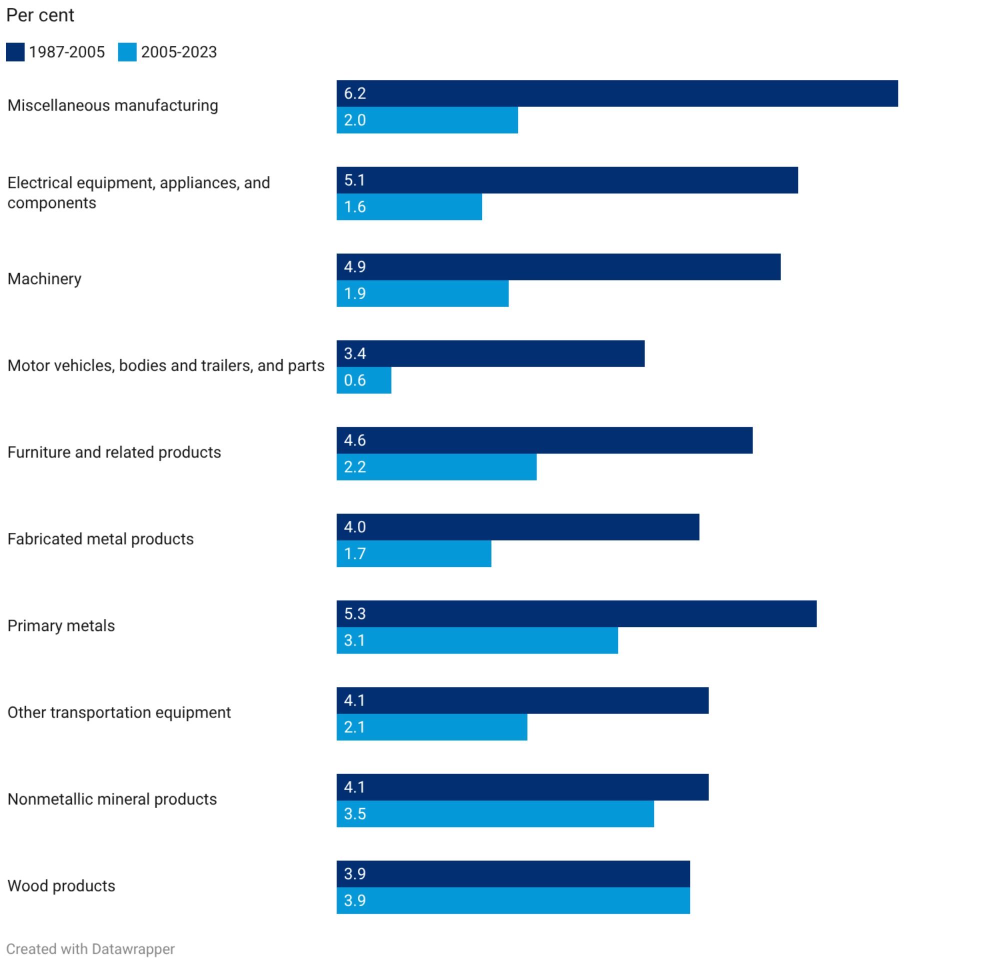
Created with Datawrapper

3. Dimensions of Difference

If the productivity growth slowdown were accounted for by just a few industries and not by the rest of the 19 three-digit industries, then we could focus on a limited set of causes unique to those industries and not to others. But that approach does not work for the U.S. manufacturing productivity growth slowdown, because every one of the 19 industries experienced a post-2005 slowdown in labour productivity growth.

The price-adjusted productivity growth rates for 1987-2005 (dark blue bars) are contrasted with 2005-23 (light blue bars) in the twin charts of Figures 3 and 4. The magnitude of the slowdown is visible as the difference between the length of the dark and light blue bars. Durable goods industries are shown in Figure 3, and the

Figure 3: Annual Average Labour Productivity Growth of Durable Manufacturing Excluding Computers, 1987-2005 vs. 2005-2023

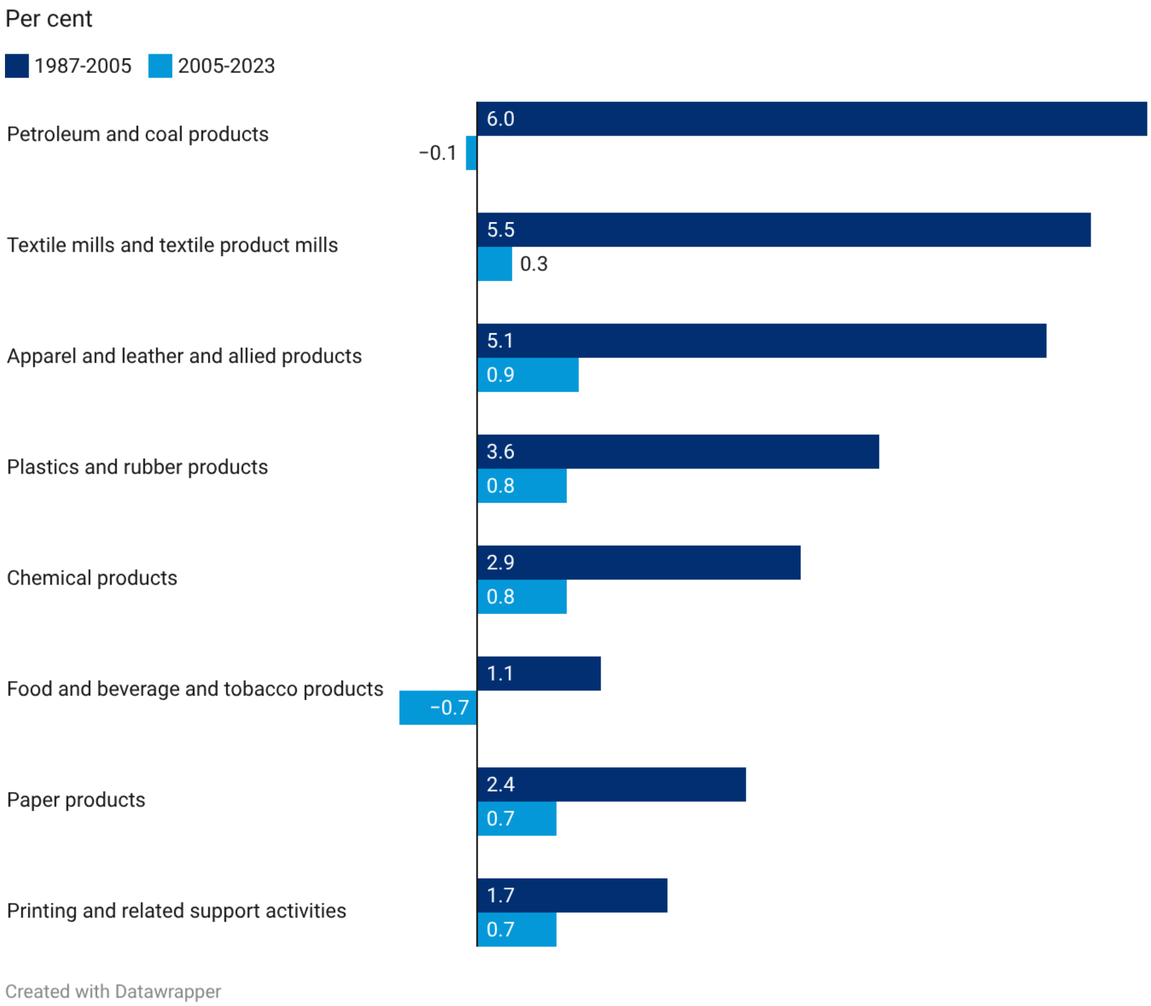


computer industry is excluded to avoid distorting the scale. The ten durable goods industries in Figure 3 are ranked in descending order of the post-2005 slowdown.

An interesting feature of the ranking in Figure 3 is that the industries at the bottom with the smallest slowdowns achieved this standing not just because their productivity growth was higher after 2005 but because it was lower before 2005. On average the bottom four industries slowed from a pre-2005 growth rate of 4.4 per cent to a post-2005 growth rate of 3.2 per cent, or a slowdown of 1.2 per cent. In contrast the four industries at the top of Figure 3 had a higher average pre-2005 growth rate of 4.9 per cent and lower post-2005 growth of 1.5 per cent, resulting in a slowdown of 3.4 per cent per year. A noticeable difference is that most of the industries at the top of Figure 3 with the largest slowdowns produce relatively complex products, including electrical equipment, machinery, and autos, while most of the industries with smaller slowdowns at the bottom of the figure produce relatively simple products, such as primary metals, nonmetallic mineral products, and wood products.

Figure 4 shows the same display for non-durable goods. Here the pattern is

Figure 4: Annual Average Labour Productivity Growth of Nondurable Manufacturing, 1987-2005 vs. 2005-2023



different, with all nondurable industries having productivity growth rates close to zero for 2005-23, so that the chart's ordering by slowdown rank is the same as the rank of 1987-2005 growth rates. The only exception is the food industry, which had very slow pre-2005 productivity growth and negative growth post-2005. For the nondurables in Figure 4 it is hard to discern any relationship between the slowdown by industry and the complexity of that industry's products. The largest slowdown was in the petroleum and coal products industry, which experienced a remarkable 6.0 per cent slowdown (between 6.0 per cent growth pre-2005 to -0.1 growth post-2005).

While production in petroleum refineries is relatively complex with a large capital investment required, the next two ranked industries – textiles and apparel – are not capital intensive nor complex. These two industries are the poster children of the import invasion, particularly apparel where the import penetration ratio had reached 97 per cent by 2018. Apparel productivity growth during 1987-2005 appears to be at a relatively healthy 5.1 per cent, but that rate was achieved only because hours disappeared at a much faster annual rate (-6.9 per cent) than output (-1.8 per cent). Apparel productivity growth declined to 0.9 per cent after 2005.

The computer and electronic products industry was omitted from Figure 3 due to its extended horizontal scale and is shown separately in Figure 5. The top four bars trace its steady decline in productivity growth from 19.4 per cent during 1987-2000 to 7.2 per cent in 2010-23. Next on the chart are the dark and light blue bars recording the decline in productivity growth from 18.7 during 1987-2005 to 9.4 per cent in 2005-23. The bottom pair of two bars show the difference made when the Byrne price adjustment is excluded and we return to the original BEA/BLS growth rates that are 6.5 and 6.3 per cent slower, respectively.

The stark 9.3 per cent post-2005 decline in productivity growth for the computer industry has rightly attracted more attention than in any other manufacturing industry. We return below to Moore's Law and its evolution over time along with other factors related to import competition that help as explanations. Here we note that, second only to the 97 per cent of the apparel industry, by 2018 the import penetration ratio for the computer and electronics industry had reached 84 per cent. The decline of this industry is inseparable from its mass offshoring migration to Asia.

4. The Import Invasion and Its Implications

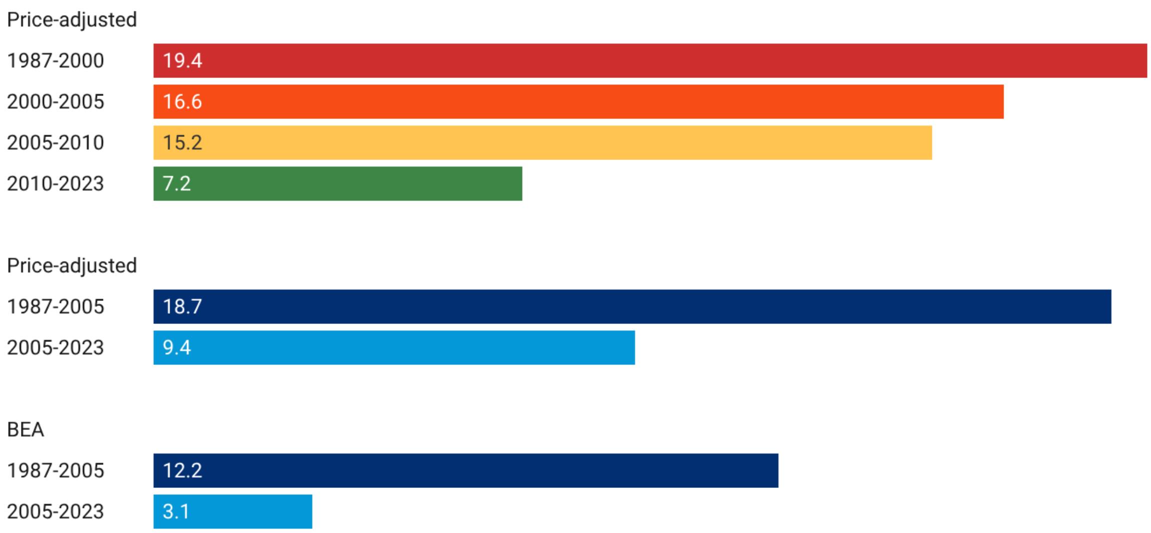
Our primary focus is on the effects of import competition starting in the late 1990s as an important explanation of the productivity growth slowdown that is usually interpreted as starting a decade later. Figure 6 illustrates the expansion of imports from China, Mexico, Canada, and the rest of the world not as a percentage of all imports, but rather as a percentage of domestic U.S. manufacturing output. The import measure encompasses both intermediate goods and final goods that physically arrive in the United States. These ratios are shown for 1991, 2000, 2010, and 2023. Using the abbreviations IM for nominal imports and GO for nominal gross output, the import ratios in Figure 6 are IR = 100*IM/GO.

Figure 6 shows that for all countries the IR increased between 2000 and 2023 from 28 to 45 per cent.[7] The decade 2000-2010, when so many domestic manufacturing jobs were destroyed, deserves the term "China Shock," in the sense that the IR for China more than tripled during that decade, growing by 5.4 per cent of U.S. manufacturing output. China alone accounted for slightly more than half of the 10.1 percentage point increase in the IR for all countries during that decade. By comparison the Mexico IR in the same decade increased by only 1.4 percentage points. By 2023, however, the Mexico IR ratio increased by another 2.4 points from 2010 while the China ratio actually decreased by

7 The IP ratio (IM/(IM+GO)) is not redundant and is useful in an industry like apparel where domestic sales consisted of 3 per cent domestic production and 97 per cent imports. In this example the IP ratio is 97 per cent whereas the IR of 97/3, or 3200 per cent, is a ratio that is intuitively meaningless.

Figure 5: Annual Average Labour Productivity Growth in the Computers Industry

Per cent

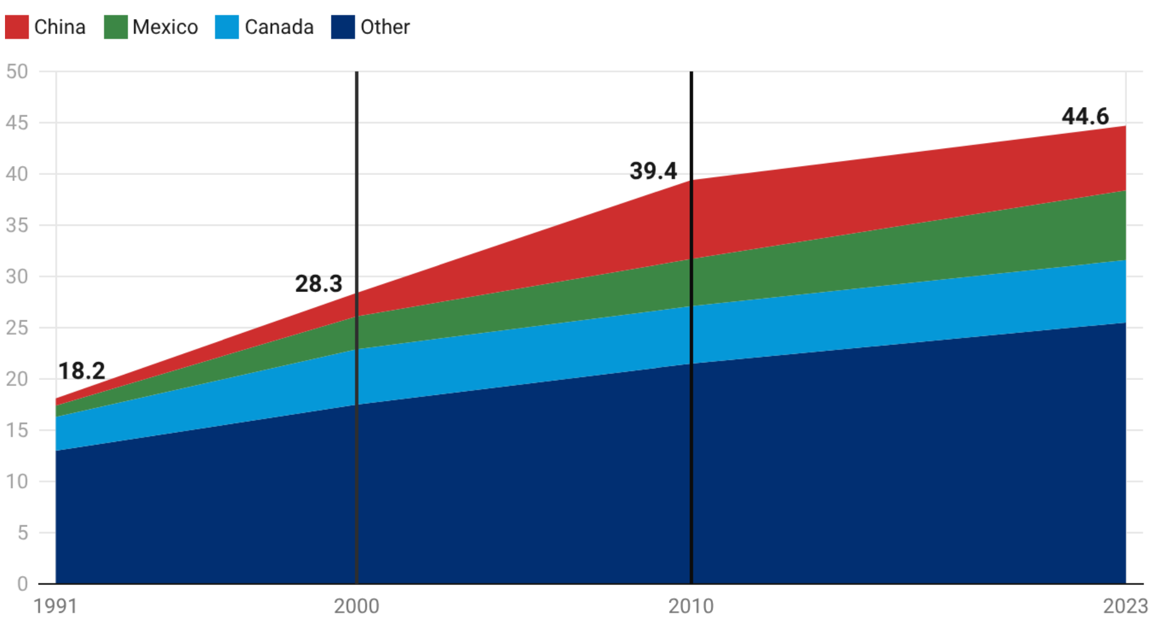


Note: The computer industry is NAICS 334.

Created with Datawrapper

Figure 6: Import Share of U.S. Manufacturing Gross Output, 1991-2023

Per cent



Created with Datawrapper

1.4 points, and Mexico achieved the distinction of becoming the top importer to the United States.

4.1 Import Invasion and Output Stagnation

We can validate the role of rising imports as a major explanation for the disappearance of output growth after the year 2000. As shown in Figure 2 above, for the 18 industries excluding computers and including the price adjustments, the annual growth rate of output declined from 3.0 per cent per year in 1987-2000 to 0.1 per cent in 2000-10. Was the transition to post-2000 output stagnation related to the import invasion? The 2000-10 growth rates of output in the 18 industries are significantly negatively correlated with the rise of the import penetration ratio in each industry. The correlation coefficient is -0.69 of the individual industry 2000-2010 output growth rates with the 1989-2005 change in the IP ratio, and this is significant at the 1 per cent level.

The import surge not only reduced output growth in the 2000-10 decade and decimated hours and employment, but it had other effects as well on investment, R&D, and the pace of innovation. The displacement of domestic demand was the most obvious of these channels. Standardized goods impacted by low-cost foreign competition were purchased by consumers looking for the lowest prices, and in industries like textiles, apparel, toys, consumer appliances, and furniture, the demand for domestic production melted away. Between 2000 and 2010 output fell by 71 per cent in the domestic apparel industry, by 50 per cent in textiles, 38 per cent in furniture, and 33 per cent in electrical equipment and appliances.[8]

Not just in these most-exposed industries but in many others, import competition exerted intense pressure on prices and profit markups. Feenstra and Weinstein (2017) document the downward pressure on markups after 2000. Declining markups and profits directly reduce the resources available for capital investment and R&D, thus creating an indirect channel of causation from expanding imports to a decline in capital deepening and innovation.

Many firms responded to this profit pressure by offshoring the supply chain to reduce costs, in addition to their cost-reducing reductions of employment. This in turn resulted in the offshoring bias, as discussed above. As we have seen this leads to the overstatement of growth in RVA and TFP. Since most of the offshoring bias occurred during the 2000-10 decade, it raised the growth rate of TFP during that decade more than after 2010. This implies that studies of the TFP growth slowdown that use a 2010 break date (like Atalay *et al.*, 2025) overstate TFP growth pre-2010 relative to post-2010 and thus exaggerate the magnitude of the TFP slowdown. This occurs in addition to the overstatement of the post-2010 slowdown caused by the 2009-10 cyclical distortion discussed above.

8 These percentages come from our data file on price-adjusted gross output for the 19 individual industries. See the Data Appendix.

4.2 Reallocation and the 2000-10 Productivity Surge

We previously noted in the context of Figures 1 and 2 above that growth in both output and hours slowed by the same amount between 1987-2000 and 2000-10, implying that productivity growth exhibits no slowdown between those two intervals. But after 2000 the manufacturing economy suddenly shifted gears from healthy output growth with stable hours, to a very different regime of stagnant output with evaporating hours. Autor *et al.* (2013) explain why the employment decline was so persistent. Many of the plants most affected by the import invasion were located in relatively small cities and towns with few alternative employment opportunities. Displaced workers often could not afford to move, in part because the closing of the local factory decimated the local housing market and evaporated home equity, preventing moves to higher-cost locations.

The post-2000 transition changed the main source of productivity growth from innovation and capital deepening to reallocation when low-productivity plants closed and the mix within firms and industries shifted to higher productivity plants and firms. This mechanism is validated both theoretically and empirically by Bernard *et al.* (2006). The gain in productivity observed in our industry data does not represent healthy innovation but rather the closing of low productivity plants and the shift in production to higher efficiency plants. These authors document the impact of trade on the closing of inefficient plants. While productivity growth, whether through rising output or falling labour input, is not a concern in and of itself, efficiencies gained from mass closures of domestic plants pose welfare concerns for affected regions. Not only did single plants close but so did entire firms in the textile, apparel, and furniture industries, as well as such iconic firms as RCA and Zenith in consumer electronics. Melitz (2003) provides additional evidence on the intra-industry reallocation effect.

Bloom, Draca, and Van Reenen (2016) found the same effect in Europe, where a similar wave of Chinese imports caused inefficient plants to close and the more efficient plants to restructure, often by shedding labour. Kim (2019) reported a similar finding with the firm-level data in the Canadian manufacturing sector that showed rising import penetration from China shifting economic activities towards high productivity firms which offset the declines of TFP within firms. Thus, both in Europe and North America productivity growth during 2000-10, which appears to provide evidence of continuing healthy innovation as before 2000, actually may have been partly or largely due to plant closings and compositional effects. Since our data are totals for each of the 19 industries without intra-industry firm detail, we cannot measure the within-industry reallocation effects. There is little cross-industry correlation between the post-2000 change in output or productivity growth and the level of 2000 output per hour, indicating that the reallocation effect occurred within industries rather than between industries.

4.3 Investment, R&D, and Innovation

Competition from imports squeezed profits not only in the most exposed industries but throughout manufacturing. The negative impact on domestic investment was augmented by uncertainty and the increasingly attractive option of offshoring production and supply chains. Autor *et al.* (2020) document that firms most exposed to imports were most likely to reduce capital expenditures. By choosing not to invest in the newest and most efficient automation and manufacturing technologies, these firms lost the chance to implement innovations that were occurring in their particular industries.

As discussed above in our summary of Figure 1, the average annual growth rate of manufacturing output (including computers and price adjustments) slowed from 4.1 per cent in 1987-2005 to 1.3 per cent in 2005-23. Over the same period the average growth rate of real investment declined by about the same amount, from 5.0 to 1.0 per cent per year. Pierce and Schott (2017) highlight the role of the post-2000 Chinese import explosion in reducing investment in the subset of industries most exposed to import competition. Offshoring of the production process also played an important role in declining investment. A large literature on trade (Bernard, Jensen, and Schott, 2006; Acemoglu *et al.*, 2016) documents that increasing imports led to a decrease in capacity utilization and lowered the necessity of additional capital investment within the U.S.

Another cause of lower investment growth in the early 2000s can be traced to the increasing financialization of the U.S. firms. Lazonick (2014) argues that corporations shifted towards a "downsize-and-distribute" regime that redirected earnings to financial interests instead of investment in production capacities, a trend that took off especially after the 2003 Securities and Exchange Commission rule change that made stock repurchases easier. Gutiérrez and Philippon (2017) show that 80 per cent of the decline in investment after 2000 can be accounted for by the amount firms spend on share buybacks.

The process by which low-cost import competition reduced investment had a similar effect on domestic R&D. Industries such as consumer electronics and primary metals shifted R&D activities abroad or concentrated them in large multinational firms that located R&D facilities in foreign countries rather than in the United States. There was also a process by which innovation shifted out of manufacturing to firms in the information technology industry like Apple, which developed software in Silicon Valley and other domestic locations while outsourcing the implementation of manufacturing process innovation to the Asian locations where the hardware device production was concentrated.

Using pre-2007 data from manufacturing patenting, Autor *et al.* (2020), show a decline in U.S. inventors' patenting between 2000 and 2007 (which is the end of their sample period) can be linked to import competition. They find that this exposure leads to a decline in private firms' R&D expenditure and a decline in patent applications. However, it is not clear that offshoring necessarily leads to fewer innovations. The authors also find that the negative impact of imports on innovation performance is larger in less profitable and less capital-intensive firms.

More broadly there are three possible reasons why offshoring may lead to a decline in manufacturing productivity growth. First, offshoring renders researchers and engineers unfamiliar with the manufacturing process, hindering further innovation. As the vice president of General Electric bluntly puts it in an interview with MIT Technology Review (2012), "you can design anything you want but if no one can manufacture it, who cares?" Consequently, offshoring is not simply moving low-skilled jobs abroad, but also makes "businesses dependent on someone else's innovation for next generation products." Leveraging an unexpected bilateral trade deal between U.S. and China in 1999, Bena and Simintzi (2025) find that offshoring reduces the willingness of firms to develop labour-saving technology for existing products.

Second, import competition also leads to a reallocation of researchers from manufacturing to service industries. Xu and Gong (2017) identify 47 science and engineering occupations. They find that for research occupations more exposed to import competition, researchers tend to shift from manufacturing to service industries (finance, personal services, business and repair services). For example, one standard deviation increase in occupation-level import competition leads to an 11.5 per cent increase in the share of researchers working in business and repair services.

Third, even if import competition reallocates research effort to firms with more market power, it does not always follow that more dominant firms become more efficient, and smaller, less efficient firms exit the market. There has been a decline in business dynamism in the US, and the dominant firms may have become less efficient over time (Gutierrez and Philippon, 2020; Covarrubias *et al.*, 2019). Decker *et al.* (2016), show that within the manufacturing sector, the contribution of labour reallocation to TFP growth has declined.

4.4 Outsourcing of the Computer Industry to Asia

As we have seen, the import penetration ratio of the computer industry (NAICS 334) in 2018 had reached 84 per cent, almost as high as low-tech apparel at 97 per cent. The departure of the U.S. industry with the most rapid rate of productivity growth is a central element in the process by which the import tsunami reduced productivity growth in U.S. manufacturing. While the offshoring of apparel, furniture, toys, and other basic consumer products was primarily due to lower labour costs abroad, the surrender of the computer industry is a more complex tale, combining technological shifts, under-investment, U.S. managerial priorities and shortsightedness, and scale effects as ever-increasing Asian production further reduced costs through economies of scale while the reverse process took place in the residual production that remained in the United States.

The primary driver of computer offshoring, unlike the case of apparel and toys, was not low labour costs. Dedrick, Kraemer, and Linden (2010) have shown that labour represents only about five per cent of the cost of manufacturing computer hardware. The offshoring of computer production bears some similarity to the auto industry, where in the 1980s Japanese firms pioneered the "just-in-time" production system that produced automobiles that were not just less expensive but were of higher quality and were more reliable.[9] Asian makers of computer components excelled in "process engineering." Firms in Japan, South Korea, and Taiwan learned to excel at continuous improvement, rapid defect detection, and tight tolerances. Additional sources of electronics offshoring to Asia include government subsidies, the geographical clustering of the supply chain, and greater support for worker training.

5. Innovation and Diminishing Returns

5.1 Declining Innovation

The diminished contribution of innovation to productivity growth can be divided into three separate causes, all involving R&D. The first is the decline of public research, the second the retreat of corporate research from basic science, and the third and perhaps most important, decreasing returns to R&D investment. This set of factors is particularly important for understanding the manufacturing slowdown, since that sector of the economy accounts for two-thirds of R&D expenditure and the related problems are more acute in manufacturing than in the rest of the economy.

Public R&D expenditure declined steadily from its peak of 2.0 per cent of GDP in 1964 to only around 0.7 per cent of GDP in recent years. Gruber and Johnson (2019) argue that revitalizing public research is crucial for future productivity growth, based on the traditional role of positive externalities. In contrast private firms often shun long term projects that are potentially beneficial to society. For instance, Pfizer terminated its R&D efforts on Alzheimer's and Parkinson's diseases in early 2018, not because of a lack of funds, but because the patent protection period was too short for the firm to make a profit. In contrast, public research has a long-term horizon. Azoulay *et al.* (2019) and others point to the army-sponsored Advanced Research Projects Agency (ARPA) which has played a substantial role in the early stages of high-profile inventions, including the internet, personal computers, lasers, and Microsoft Windows.

The decline in public research was aggravated by the retreat of corporate research. Arora *et al.* (2019), emphasize that U.S. research prior to the 1980s was characterized by giant corporate labs, such as the Bell Lab of AT&T, and research units of DuPont and Xerox, all of which were manufacturing firms. These corporate research clusters focused on general purpose technologies and worked across disciplines on a large scale. However, large companies shifted from general basic science to

9 Consumer Reports magazine in its 2025 automobile quality and reliability rankings awards the top 5 places to Japanese and Korean brands, with cars made by American firms perennially relegated to the bottom of the list.

narrow commercial development starting in the 1980s. As Arora *et al.* (2015) show, companies became less willing to invest in basic science in part because the results could benefit business rivals.

In recent years attention has shifted to the third factor, diminishing returns to research investment. This in part reflects the influence of a much-cited paper by Bloom *et al.* (2020), claiming that "new ideas are getting harder to find." Their two main examples are the development of new drugs by pharmaceutical companies and, for computers, Moore's Law showing that the number of transistors on a semiconductor chip doubles every two years. The steady exponential growth embodied in Moore's law has required an increase in the number of research workers by a factor of 18 over the previous four decades.

5.2 Fading Pharmaceuticals

Our 19-industry database covers three-digit industries including chemicals (NAICS 325) but not four-digit industries including pharmaceuticals (industry 3254). BLS (2018) shows that this four-digit industry makes the fourth-largest contribution to the manufacturing TFP slowdown from 1992-2004 to 2004-2016. Growing R&D costs and a higher termination rate of projects are two prominent corollaries of the slowdown. Anticipating in part the Bloom *et al.* (2020), research, Deloitte (2018) estimates that the average R&D costs of developing a compound from discovery to launch almost doubled from 2010 to 2018.

The slowdown in pharmaceutical innovation may have started much earlier. Gordon (2016) notes that the decades between 1940 and 1970 witnessed the invention or the widespread usage of many important drugs and medical techniques (such as an array of antibiotics, computed tomographic imaging, polio vaccine, and many others), but the rate of innovation slowed down in the decades that followed. Bloom *et al.* (2020), provide empirical evidence that the research productivity in medical research, defined as the ratio of years of life saved to the number of publications, first increased from 1975 to the mid-1980s and then fell. These authors measure the average annual growth rate of research productivity to be -0.6 per cent for all cancers, -6.8 per cent for breast cancers, and -3.7 per cent for heart disease. The same authors show that lives saved per million in clinical trials per real dollar of research expenditure fell over the same period by a factor of eight for breast cancer research and by a factor of 16 for all cancer research. More recent research has christened "Eroom's Law" (Eroom = Moore spelled backwards); this shows that drug approvals per billion dollars of real research expenditures declines by half every nine years. Taking the results for computers and pharmaceuticals together, they imply that 60 per cent of total manufacturing R&D expenditure is suffering from diminishing returns.

A straightforward cause of this phenomenon is that it has become increasingly challenging to improve the understanding of basic science. The lab-based process of discovering new drugs and compounds makes it challenging to predict the behavior of materials and requires multiple lab experiments which are costly, time-consuming, and unpredictable. Further, chemical companies face an increasing number of long-term disruptions in the form of more foreign competition, rapid shifts in end-market demand and a growing burden of environmental regulation.

5.3 Automobile Recalls Reduce Productivity

A higher frequency of car recalls since 2012 has reduced industry productivity growth. The cause of increasing recalls may be the increasing complexity of vehicles (Harbour *et al.*, 2015). In addition, researchers at McKinsey conjecture that since many companies now have common product platforms and supply-chain partners, one defect on a single module can negatively affect multiple vehicle models (Aragon *et al.*, 2019).

Auto recall services, if they cannot be carried out by dealer service departments, negatively affect car manufacturers' productivity, since the factories have to make replacement parts, which raises labour costs and labour input for a given number of vehicles produced. Moreover, car recalls can be costly in themselves. For example, in 2016, GM recalled 23 million vehicles in the U.S. which cost GM $4.1 billion; in 2015, automakers and their suppliers together paid $17.5 billion on claims and warranty accruals (Automotive News, 2018). This issue relates to our previous comments about the persistent low reliability scores of automobiles produced by American-owned firms as tallied by Consumer Reports.

5.4 Productivity Shrinkage in the Food Industry

As shown in Figure 4 above, the food and beverage industry has recorded negative -1.5 per cent annual productivity growth since 2005. Day-Rubenstein and Fuglie (2012) argue that in recent years, new product development has been driven by consumer demand, as opposed to reducing costs or reducing resources needed for production. "An estimated 20,000 new food products are introduced in the U.S. annually. While some of these new products embody technical change, only about 10 per cent are thought to be true innovations. The average lifespan of a new food product is relatively short."

6. A Catalog of Causes

In addition to resulting from the import invasion and a slowing of innovation, the productivity growth slowdown in U.S. manufacturing has additional causes. Further contributing factors include (1) the failure of robotics to boost productivity growth in the industries where the population of robots has expanded rapidly, (2) environmental and other government regulations, and (3) a persistent shortage of skilled labour. In this section we identify these causes and provide examples for the six industries that make the highest numerical contribution to the post-2005 productivity growth slowdown: computers, petroleum, pharmaceuticals, autos, machinery, and food. Together these six industries explain 71 per cent of the post-2005 slowdown.

6.1 Robotics and Automation

The use of robots expanded rapidly in the past 15 years in some manufacturing industries but did not prevent them from recording zero or minimal productivity growth. Why did the increase in the adoption of robotic technology fail to boost productivity growth? The first answer is the most important. Despite its continued growth, robotics accounted for only 1.1 per cent of total equipment investment expenditure in the manufacturing sector in 2021 (Annual Capital Expenditure Survey). Thus, the gains from robotics automation were swamped by other factors such as lower output demand or falling utilization or required responses to regulations, all of which tended to mask the effect of added robots.

What are the other reasons why robots have had a disappointing effect on manufacturing productivity growth? Based on the task-based model of Acemoglu and Restrepo (2019), numerous articles have shown that increased robot use raises labour productivity. Graetz and Michaels (2018) find that increased adoption of robots from 1993 to 2007 contributed 0.36 percentage points to annual labour productivity growth using panel data of seventeen countries. Acemoglu and Restrepo (2020) use the same data but focus on the U.S. labour market and find that one more robot per thousand workers reduces the employment-to-population ratio by 0.2 percentage points, which seems to match the narrative of the task-based model where automation displaces low- to middle-skill labour.

On the other hand, Benmelech and Zator (2025) find using cross-country and German administrative data that firms invest in robots when they face difficulties in finding workers and subsequently increase employment after the investment. This limits the economic impact of robots. Moreover, a significant portion of the increased robotics use is concentrated in the auto industry, which has mainly invested in robotics to drive its electric vehicle transition and respond to labour shortages (IFR, 2024). Given that the auto industry was responsible for one-third of the annual installations of industrial robots in 2024, it may be that other issues that reduce auto-industry productivity growth have offset the benefits that robots would be expected to provide.

A key challenge in automation is to ensure a smooth integration of new technologies to existing production facilities. Knoess *et al.* (2017), highlight that automation and robotization do not necessarily lead to productivity gains for all firms. Over the past two decades, the leading auto factories only automate "the simplest, most repetitive processes" such as the paint and body shops where they see the greatest gains from automation.

Skill shortages, which we discuss below, could partly be responsible for the unimpressive productivity impact of industrial robots in recent years. Between 2015 and 2024, the robot density defined as the number of industrial robots per 10,000 workers in the manufacturing sector almost doubled in the United States from 176 to 307, whereas China with its excess supply of STEM graduates saw a ten-fold increase from 51 to 567 within the same period (Müller, 2025). The lack of skilled workers who can manage these robots could be exerting downward pressure on the U.S. robot adoption rate.

6.2 Regulatory Burdens

Regulation changes that may impact productivity growth include tighter environmental laws, stricter pollution controls, and added safety standards. Heavy manufacturing industries such as petroleum, chemicals, and motor vehicles were most exposed to regulations that redirected capital investment or increased compliance costs. The food and beverage industry also faced tighter controls because of increased concern for public health and food safety in recent decades. In contrast the machinery and computer industries were largely unaffected by changes in regulations.

The Clean Air Act Amendment of 1990 called for the reformulation of motor fuels to reduce emissions from motor vehicles, which required firms to process gasoline with a higher percentage of oxygen. Between the late 1990s and early 2000s, EPA significantly expanded enforcement of regulations that discouraged firms from investing in new refineries or modifications to existing refineries as needed for capacity growth. The New Source Review (NSR) program, for example, required permitting before construction or modification of major stationary sources, including oil refineries, to ensure installation of pollution control equipment. NSR enforcement and interpretive uncertainty imposed both delays and cost increases on refinery projects, discouraging capacity expansion and adding permitting burden (Senate Hearing 107-868).

The Food Safety Modernization Act in 2011 was a significant shock to the regulatory framework for the food and beverage industry that changed the regime from responding to food contamination to requiring preventive measures by manufacturers. These included Hazard Analysis and Risk-Based Preventive Controls (HARPC) plans and increased inspection frequency and extra regulatory authority by the FDA. Firms responded to this change by increasing non-production labour for quality control, documentation of plans, and compliance, which would have minimal impact on output expansion. Our price-adjusted data record an output growth rate for the food industry of just 0.2 per cent per year between 2010 and 2023 while capital input increased by 2.1 per cent per year and investment grew by 3.2 per cent. This is consistent with a redirection of investment from the achievement of output and productivity growth to the compliance with regulatory requirements.

The main channel of regulatory consequences on labour productivity of the motor vehicle industry has occurred through changes in the fuel economy and emission standards. The change in the fuel economy standards imposed a binding constraint on auto manufacturers and forced them to enact engineering changes rather than complying through controlling the product mix of passenger cars and trucks. Moreover, added pollution controls required advanced catalytic converters and engine controls which increased vehicle cost and complexity and capital investment without much output effect (Wang and Miao, 2021).

The chemical industry was significantly impacted by the set of regulatory changes after 2000. Firms were required to invest heavily in end-of-pipe abatement, scrubbers, and monitoring systems. A series of papers already studied the effect of environmental regulation on manufacturing productivity. Greenstone, List, and Syverson (2012) use the 1972-1993 plant-level Census of Manufactures data to show that plants located in the non-attainment areas that faced much stricter regulation exhibited lower total factor productivity growth rates than the plants in the attainment areas. They estimate that the organic chemicals industry experienced a TFP decline of roughly 17 per cent. Shapiro and Walker (2018) show that increased regulatory burden on the manufacturing sector through pollution taxes led to abatement investment rather than productivity-enhancing technology development.

6.3 Skill Shortages

The BLS reported for 2025Q3 that more than 400,000 positions in the manufacturing sector remained open and were not filled. A survey by the National Association of Manufacturers (NAM) shows that attracting and retaining talent has been ranked as the primary business challenge consistently since 2017 (Deloitte, 2024). This shortage of skilled workers in manufacturing is not a new phenomenon. Instead, it has developed over the past two decades through the interplay of structural changes in the labour force, educational preferences biased against manufacturing jobs, and the evolving economic environment and advances in technologies.

The skill gap can be dated back to 2001, when 80 per cent of manufacturers reported a moderate to serious shortage of qualified job applicants in a survey by the NAM. This was followed by the earliest wave of retirement of skilled workers in the 2000s who entered the workforce in the hiring boom of the 1970s and 1980s (National Academies of Sciences, 2017). Their exit meant losses in tacit knowledge and skills accumulated over decades, but U.S. firms increasingly stopped training workers internally and shifted the burden of preparing occupational skill and knowledge to schools and students, which widened the skill gap even more (Cappelli, 2015). The Great Recession of 2008-09 also contributed to this trend by permanently altering the skill pipeline, as the entry-level positions disappeared and did not fully recover (Mullins and Forbes, 2015).

In part because of the aging and retirement of the skilled workforce, the supply of qualified workers stagnated. From 2011 to 2022, there was no growth in the number of associate degrees, which prepare graduates for high-skilled trades. More high school students chose to go to college instead of trade schools, as the educational system placed more weight on college attendance. The impression of manufacturing jobs as dangerous or dirty further discouraged students, though these perceptions were often misconceived or outdated (Stockman, 2025). While the supply of skilled workers has stagnated, the demand has continued to expand in response to recent advances in technology and the complexity of machinery. Overall, the difficulty of finding high-skilled workers in recent decades suggests bottlenecks and complications in production processes which contributed to the productivity growth slowdown across so many of the component industries within manufacturing.

The auto industry has been hit hardest in terms of the shortages of skilled labour, especially positions such as industrial maintenance technicians, automation and controls machinists, and tool and die makers. These occupations are difficult to fill as they require a long period of training and combine knowledge from multiple disciplines in mechanical, electrical, and software engineering. With the skill pipeline broken after the Great Recession, manufacturers in the auto industry have consistently reported difficulty in obtaining high-skilled workers for their factories. More recently, another source of skill shortage emerged from a geographic mismatch between areas where internal combustion engine vehicles manufacturing jobs are being lost and where new battery electrical vehicles manufacturing jobs are emerging, which highlights multi-faceted struggles in the search of skilled labour (Saha *et al.*, 2025). Both petroleum and chemicals sectors rely on high-skilled plant operators who facilitate continuous-flow production, which exposes them to the skill shortage problem due to decreases in apprenticeship programs and the retirement of experienced workers. Moreover, the two industries are heavily capital-intensive and safety-constrained, hence the loss of skills in these industries often shows up as lower utilization, not lower employment. The tightened regulations and safety standards in recent years also contributed to the high demand for instrumentation and control technicians who are essential to ensure safe operation under the constraints imposed by environmental rules.

7. Conclusion

Official BEA/BLS measures show that U.S. manufacturing labour productivity growth disappeared after 2010, collapsing from +3.3 per cent per year during 1987-2010 to -0.3 per cent per year between 2010 and 2023, representing a slowdown of 3.6 per cent per year. How could a sector that had propelled American growth since the founding of the Republic stumble so badly into apparent stasis? An emerging literature has attempted to explain this slowdown by changes that occurred after 2010.

This article starts the story ten years earlier. Over the decade 2000-2010 manufacturing employment fell by 44 per cent from 17.3 to 11.5 million. The wave of imports that caused this employment evaporation was christened by David Autor and his co-authors as the "China shock", but we refrain from using this term because over the longer period 1991 to 2023 China accounted for only 22 per cent of the increase in the ratio of imports to domestic manufacturing output. Reflecting the fact that nations worldwide have increased the penetration of imports into the U.S., we prefer the term "import invasion."

Just as stark as the disappearance of productivity growth after 2010 was the disappearance of output growth ten years earlier. As one U.S. industry after another witnessed its sales melt in the face of import competition, the response was not just to cut employment and close plants but also to experience a squeeze on profits, a decline in capacity utilization, a decline in investment both in fixed capital and R&D, and indirectly a decline in the pace of innovation. In many cases suppliers of components moved offshore before producers of final goods, leading to emerging gaps in the domestic supply chain. While measured productivity growth during the 2000-2010 decade was the same as during 1987-2000, it occurred in a completely different industry environment. The same productivity growth that in 1987-2000 was the difference between healthy growth in output and zero growth in hours of work, instead in the 2000-10 decade was the same arithmetic difference between stagnant output and evaporating hours.

Our measure of the productivity growth slowdown makes three changes from most of the literature that uses BEA/BLS data to assess the slowdown that begins in 2010. We adjust the deflators of the BEA gross output data by incorporating into our output measures the deflator adjustments proposed by Atalay *et al.* (2025), to improve the treatment of quality change in existing deflators. These boost the growth rate of output and productivity in total manufacturing and a few durable goods industries, with little effect for other durables and most nondurable goods industries. The switch to these deflators does not help explain the slowdown but rather increases the puzzle, as the price adjustments boost the post-2010 manufacturing productivity slowdown from -3.6 per cent to -4.1 per cent per year.

Our second change is to switch the break date for the slowdown from 2010 to 2005, due to the unusual positive bubble of productivity growth in 2009-10 resulting from recession-caused shedding of labour. This switch in the break year reduces the magnitude of the slowdown from -4.1 per cent when dated as starting in 2010 to -3.5 per cent when the break date is 2005. We join other authors in pointing to the computer and electronic products industry as responsible for a disproportionate share of the slowdown. When measured for 18 of the 19 manufacturing industries, excluding computers, the slowdown declines from -3.5 to -2.5 per cent per year. Thus, we eliminate 39 per cent of the slowdown ((4.1-2.5)/4.1) prior to the start of our substantive analysis. While this might seem to be defining away the problem, the sources of the computer industry slowdown are better understood than for most of the remaining 18 manufacturing industries.

Another difference with some past research is that we focus on labour productivity rather than TFP. A measurement issue called "offshoring bias" leads to an understatement of growth in intermediate inputs. Since value added growth is obtained by subtracting from gross output growth the understated growth in these inputs, growth in both value added and TFP is overstated. Since the offshoring bias was most significant when imports were expanding most rapidly, this implies that the growth in TFP is overstated more before 2010 than afterwards, leading to an exaggeration of the TFP slowdown when 2010 is the break year for defining the slowdown. This source of overstatement of the slowdown is avoided in this article by limiting our attention to labour productivity growth, defined as the growth in gross output per hour.

We examine differences in the post-2005 slowdown across the 19 three-digit manufacturing industries, ranging from -12 per cent per year for computers to -0.2 per cent for wood products. For durable goods the extent of the slowdown is greatest for complex products like computers, autos, electrical equipment, and machinery, and smaller for less complex products like furniture, wood, and nonmetallic minerals. No such pattern emerges for nondurable goods, where the largest slowdowns are for the most capital-intensive industry – petroleum – and the two least capital intensive – textiles and apparel. The fact that 14 of the 19 industries experienced a post-2005 slowdown of more than -2 per cent suggests that the underlying causes of the slowdown are largely common to many industries rather than specific to each industry.

Our discussion of the import invasion establishes a causal chain between the import invasion and the post-2000 disappearance of output growth. Across the 18 industries excluding computers there is a highly significant negative correlation between the post-2000 growth of industry output and the 1987-2005 growth of that industry's import penetration ratio. We trace a channel of causation from the arrival of imports to declining domestic sales and employment, lower profits and capacity utilization, and less investment in fixed capital and R&D.

The effect of imports on domestic innovation is more complex. Some authors point to a shift of innovation activity toward larger leading firms both inside and outside of manufacturing, in which case innovation shifts its location rather than experiencing a decline. Others argue that domestic manufacturing and innovation are complements; when production of components is offshored the growing distance from the production process inhibits further improvements that combine design and process innovation.

The computer and electronic products industry, which contributed most to manufacturing productivity growth slowdown, largely offshored production to Asia after 2005. By 2018 its import penetration ratio had reached 84 per cent. This did not occur primarily because of lower labour costs, as labour makes up only five per cent or less of manufacturing cost for most electronic products. Instead, the attraction of Asia was its emphasis on process innovation. "Manufacturing optimization" leads to the ability to ramp up production quickly at massive scale. Further explanations include government subsidies, geographically concentrated supplier clusters, and support for worker training.

The article then turns to factors that help to account for the declining contribution of innovation to productivity growth. These include a reduction in public support for R&D, and a shift in emphasis of private R&D from basic science and process improvement to brand extensions and product copying. A third factor is diminishing returns to research effort as demonstrated in recent research that "new ideas are getting harder to find." Even though Moore's Law in the computer industry may continue its pace to some degree, the number of research workers needed to maintain that pace has increased multifold in the last four decades. Similarly, a given number of research workers in the pharmaceutical industry produces ever fewer patents, drug approvals, and lives saved.

We emphasize the interplay between investment and regulation. Import competition reduced profits and investment directly. Beyond that regulations and other structural changes diverted investment from improving productivity to adhering to regulatory demands. Auto makers had to divert resources to raising fuel economy and convert to electric vehicles, drug makers had to wait longer for FDA approvals, the chemicals industry faced safety and anti-pollution regulations, the food industry grappled with new safety regulations, and petroleum refineries had to retool to process newly developed fracking supplies and ethanol requirements.

A separate issue involving investment was the apparent failure of the rapidly expanding robot population to boost productivity growth in the auto industry, where robots are most heavily used. One explanation is straightforward —- robot investment constitutes only one to two per cent of total manufacturing equipment investment. Another involves diminishing returns — robots are already widely used in auto body and paint shops but cannot yet do sensitive assembly work requiring human hands and dexterity, particularly as electronic controls and devices make autos ever more complex.

Our introduction to this article cited increasing concern at delays and defaults in American industry more generally, from military hardware to infrastructure projects like high-speed rail. Our analysis identifies a combination of failings of government and private industry. The government has retreated from basic research and until recently has abandoned industrial policy that might have created a coordinated effort to match Asian excellence in process innovation. Private corporate research has switched from basic science to duplicative product extensions. Investment in automation and productivity-enhancing capital has been partly set aside by share buybacks and short-term profit maximization. Government and private firms have jointly failed to anticipate a growing shortage of skilled labour and have set up a comprehensive set of training and apprenticeship programs.

Finally, we return to the import invasion. To date, the literature has focused on the post-2010 decline in TFP growth. Equally important was the post-2000 disappearance of output growth. Further study is warranted about intra-industry reallocation from closed plants to more productive plants and firms to explain at least in part why productivity growth continued during the 2000-10 decade, despite the import invasion that eroded the competitiveness of American manufacturing during the same decade. The many issues addressed here support skeptics who doubt that a broad-based regime of tariffs can revive American manufacturing output, employment, and productivity growth. It may already be too late.

References

Acemoglu D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016) "Import Competition and the Great US Employment Sag of the 2000s," *Journal of Labor Economics*, Vol. 34, No. S1, pp. S141–S198.

Acemoglu D. and P. Restrepo (2019) "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, Vol. 33, No. 2, pp. 3–30.

Acemoglu D. and P. Restrepo (2020) "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, Vol. 128, No. 6, pp. 2188–2244.

Aragon A., U. Huber, T. Möller, and N. Heiko (2019) "Preventing Car Recalls with a Quality Focus," McKinsey.

Arora A., S. Belenzon, and A. Patacconi (2017) "The Decline of Science in Corporate R&D," *Strategic Management Journal*, Vol. 39, No. 1, pp. 3–32.

Arora A., S. Belenzon, A. Patacconi, and J. Suh (2019) "The Changing Structure of American Innovation: Some Cautionary Remarks for Economic Growth," in *Innovation Policy and the Economy*, Vol. 20.

Atalay E., A. Hortaçsu, N. Kimmel, and C. Syverson (2025) "Why Is Manufacturing Productivity Growth So Low?" NBER Working Paper No. 34264.

Automotive News (2018) "Auto Recall Bill Grew 26% to $22 Billion in 2016, Study Says."

Autor D. H., D. Dorn, and G. H. Hanson (2013) "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, Vol. 103, No. 6, pp. 2121–2168.

Autor D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020) "Foreign Competition and Domestic Innovation: Evidence from US Patents," *American Economic Review: Insights*, Vol. 2, No. 3, pp. 357–374.

Azoulay P., E. Fuchs, A. P. Goldstein, and M. Kearney (2019) "Funding Breakthrough Research: Promises and Challenges of the ARPA Model," *Innovation Policy and the Economy*, Vol. 19, pp. 69–96.

Bena J. and E. Simintzi (2025) "Machines Could Not Compete with Chinese Labor," *Review of Finance*, Vol. 29, No. 6, pp. 1619–1661.

Benmelech E. and M. Zator (2025) "Robots and Firm Investment," *Journal of Financial Economics*, Vol. 174, Article 104183.

Bernard A. B., J. B. Jensen, and P. K. Schott (2006) "Survival of the Best Fit," *Journal of International Economics*, Vol. 68, No. 1, pp. 219–237.

Bloom N., C. I. Jones, J. Van Reenen, and M. Webb (2020) "Are Ideas Getting Harder to Find?" *American Economic Review*, Vol. 110, No. 4, pp. 1104–1144.

Bloom N., M. Draca, and J. Van Reenen (2016) "Trade Induced Technical Change?" *Review of Economic Studies*, Vol. 83, No. 1, pp. 87–117.

Byrne D. M., J. G. Fernald, and M. B. Reinsdorf (2016) "Does the United States Have a Productivity Slowdown or a Measurement Problem?" *Brookings Papers on Economic Activity*, Vol. 1, pp. 109–182.

Byrne D. M., A. Hamins-Puertolas, and M. M. Harnish (2023) "Transistors All the Way Down: Viability of Direct Volume Measurement (and Price Indexes) for Semiconductors," Working Paper.

Cappelli P. H. (2015) "Skill Gaps, Skill Shortages, and Skill Mismatches," *ILR Review*, Vol. 68, No. 2, pp. 251–290.

Chittoor R., P. Pietraru, and A. Sharpe (2025) "The Stylized Facts of the United States Manufacturing Productivity Slowdown," CSLS Research Report.

Covarrubias M., G. Gutiérrez, and T. Philippon (2019) "From Good to Bad Concentration?" in *NBER Macroeconomics Annual*, Vol. 34.

Day-Rubenstein K. and K. Fuglie (2012) "R&D and Productivity Lag in Food Manufacturing," USDA Economic Research Service.

Decker R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016) "Declining Business Dynamism," Hutchins Center Working Paper No. 23.

Dedrick J., K. L. Kraemer, and G. Linden (2010) "Who Profits from Innovation in Global Value Chains?" *Industrial and Corporate Change*, Vol. 19, No. 1, pp. 81–116.

Deloitte and The Manufacturing Institute (2024) Talent Study.

Feenstra R. C. and D. E. Weinstein (2017) "Globalization, Markups, and US Welfare," *Journal of Political Economy*, Vol. 125, No. 4, pp. 1040–1074.

Fernald J. G. (2015) "Productivity and Potential Output Before, During, and After the Great Recession," in *NBER Macroeconomics Annual*, Vol. 29, No. 1, pp. 1–51.

Food and Drug Administration (2015) "FSMA Final Rulemaking for Current Good Manufacturing Practice."

Gordon R. J. (1990) *The Measurement of Durable Goods Prices*. Chicago: University of Chicago Press.

Gordon R. J. (2016) *The Rise and Fall of American Growth*. Princeton: Princeton University Press.

Gordon R. J. and Sayed (2024) "New Perspectives on Productivity Dynamics in the U.S. Private Economy, 1950–2023," CEPR Discussion Paper No. 19569.

Graetz G. and G. Michaels (2018) "Robots at Work," *Review of Economics and Statistics*, Vol. 100, No. 5, pp. 753–768.

Greenstone M., J. A. List, and C. Syverson (2012) "The Effects of Environmental Regulation on Manufacturing Productivity," MIT Working Paper.

Gruber J. and S. Johnson (2019) *Jump-Starting America*.

Gutiérrez G. and T. Philippon (2017) "Investment-less Growth," *Brookings Papers on Economic Activity*, Vol. 48, No. 2, pp. 89–190.

Gutiérrez G. and T. Philippon (2020) "Some Facts about Dominant Firms," NBER Working Paper No. 27985.

Harbour R., M. Hill, and J. Schmidt (2015) "The Harbour Report," Oliver Wyman.

Houseman S., C. Kurz, P. Lengermann, and B. Mandel (2011) "Offshoring Bias in US Manufacturing," *Journal of Economic Perspectives*, Vol. 25, No. 2, pp. 111–132.

International Federation of Robotics (2024) "U.S. Companies Invest Heavily in Robots."

Ji K., L. Nauta, and J. Powell (2023) "Mapping Global Supply Chains: The Case of Semiconductors," RaboResearch.

Jibrell A. (2018) "Auto Recall Bill Grew 26% to $22 Billion in 2016," *Automotive News*.

Kim M. (2019) "Does Import Competition Reduce Domestic Innovation?" *International Productivity Monitor*, No. 37, pp. 72–95.

Knoess C., R. Harbour, and S. Scemama (2017) "Prepare Your Workforce for the Automation Age," *Harvard Business Review*.

Lazonick W. (2014) "Profits without Prosperity," *Harvard Business Review*, Vol. 92, No. 9, pp. 46–55.

Melitz M. J. (2003) "The Impact of Trade on Intra-Industry Reallocations," *Econometrica*, Vol. 71, No. 6, pp. 1695–1725.

Mullins J. and B. E. Forbes (2015) "CES Employment Recovers in 2014," *Monthly Labor Review*.

National Academies of Sciences (2017) *Building America's Skilled Technical Workforce*.

New York Times Editorial Board (2025) "America Can't Make What It Needs for Its Defense," *New York Times*, December 14.

Pierce J. R. and P. K. Schott (2017) "Investment Responses to Trade Liberalization," NBER Working Paper No. 24071.

Saha D., R. Shrestha, N. Hunt, and E. Kim (2025) "How US Auto Workers Can Thrive in the Transition to Battery Electric Vehicles," World Resources Institute.

Shapiro J. S. and R. Walker (2018) "Why Is Pollution from US Manufacturing Declining?" *American Economic Review*, Vol. 108, No. 12, pp. 3814–3854.

Schott P. K. (2008) "The Relative Sophistication of Chinese Exports," *Economic Policy*, Vol. 23, No. 53, pp. 6–49.

Stockman F. (2025) "Why Factories Are Having Trouble Filling Jobs," *New York Times*.

Terry C. and N. Lesser (2018) "Unlocking R&D Productivity," Deloitte Centre for Health Solutions.

U.S. Census Bureau (2021) "Annual Capital Expenditures for Robotic Equipment."

U.S. Congress (2002) Senate Hearings on New Source Review Policy.

Xu R. and K. Gong (2017) "Does Import Competition Induce R&D Reallocation?" IMF Working Paper.

Wang Y. and Q. Miao (2021) "The Impact of Corporate Average Fuel Economy Standards," *Resource and Energy Economics*, Vol. 63, Article 101211.

The Anatomy of the U.S. Manufacturing Productivity Slowdown: Evidence from Firms and Industries

Danial Lashkari and Jeremy Pearce*

IPM Research Article

Abstract

U.S. manufacturing labour productivity growth fell from roughly 3.5 per cent per year over 1987–2007 to near zero over 2010–2022. Growth in total factor productivity (TFP) also fell to near zero over the same period. This article examines the sources of that slowdown by decomposing manufacturing productivity growth into contributions from leader and follower firms within frontier and laggard industries. We find that the slowdown is broad-based: both for labour productivity and TFP, productivity growth declined among both leaders and followers, across alternative weighting methods, and multiple industry groupings. Standard models of economic growth treat research and development (R&D) as the primary channel through which firms generate productivity growth. The broad-based nature of the slowdown raises the question of whether the translation of R&D expenditures into productivity gains has weakened. We estimate an R&D production function at both the industry and firm levels and find that the elasticity of productivity with respect to R&D is consistently larger in the earlier period than in the full sample, even as R&D expenditure has risen across firms and industries. These results suggest that the slowdown reflects declining research productivity rather than reduced innovation effort.

*Danial Lashkari is an Economic Research Advisor and Jeremy Pearce is a Research Economist at the Federal Reserve Bank of New York. Jeremy Pearce is the corresponding author (email: Jeremy.Pearce@ny.frb.org). We thank Andrew Sharpe, Paul Scheyer, Bart van Ark, and an anonymous reviewer for helpful comments. The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Rohan Bhargava and Isaiah Dalzell provided excellent research assistance.

1. Introduction

Historically, the U.S. manufacturing sector has been a pivotal driver of aggregate productivity growth, with labour productivity, defined as output per hour, growing at over 2 per cent per year through most of the twentieth century. More recently, however, manufacturing has experienced a pronounced productivity slowdown. Labour productivity fell from roughly 3.5 per cent growth per year over 1987–2007 to near zero over 2010–2022. Growth in total factor productivity (TFP, which is the residual output after accounting for labour, capital, and intermediate inputs), fell from roughly 1.3 per cent to near zero over the same periods. This slowdown is part of a broader deceleration in aggregate labour productivity across the U.S. economy (Fernald, 2015; Syverson, 2017; Byrne *et al.*, 2016; Sharpe and Chittoor, 2025; Atalay *et al.*, 2025) and is particularly puzzling given that the manufacturing sector still accounts for the majority of private-sector research and development (R&D) expenditures (Lashkari and Pearce, 2024).

The existing literature points to several possible sources. Industry-level analyses emphasize the role of leading sectors, in particular Computer and electronic products (North American Industry Classification System, or NAICS, 334), whose rapid productivity growth in the late 1990s and early 2000s has collapsed after the Great Recession (Atalay *et al.*, 2025; Syverson, 2017). Firm-level accounts, including those that emphasize divergence between frontier firms and others, suggest that the slowdown should be concentrated in follower firms that are falling behind. In many of those theories, the frontier firms continue to advance (Andrews *et al.*, 2015; Aghion *et al.*, 2023; Andrews *et al.*, 2019; Akcigit and Ates, 2023; Olmstead-Rumsey, 2022). This "best-versus-rest" narrative has become an influential framing of the productivity slowdown, linking it to rising dispersion between frontier and laggard firms. In contrast, the competing "ideas getting harder to find" hypothesis implies a broad-based slowdown affecting all firms and industries, perhaps roughly in proportion to their research intensity (Bloom *et al.*, 2020).

Two natural questions emerge from this literature. First, is the slowdown broad-based across categories of firms and industries, or is it concentrated in particular groups? Second, how is the slowdown reflected in R&D activity: is R&D spending itself declining, or is R&D becoming less effective at generating productivity growth?

This article addresses both questions by extending our previous work along several dimensions. In Lashkari and Pearce (2026), we introduce a decomposition framework that links firm-level information from Compustat to industry-level aggregates from the U.S. Bureau of Labor Statistics (BLS) and study the trends in R&D intensity and productivity. The present article extends that analysis in three ways. First, we broaden the industry split beyond NAICS 334 to a wider set of frontier industries. Second, we broaden the definitions of leader firms to include both revenue and productivity definitions of leadership. Third, we move beyond documenting the R&D–productivity disconnect and directly estimate the changes in the relationship between R&D spending and subsequent productivity growth at both the industry and firm levels.

Our first finding is that the slowdown is broad-based across frontier and lagging industries and across firms. When leader firms within each industry are defined by employment or revenue, which are both relatively persistent measures of firm size, both leaders and followers slow down. This result holds across different weighting methods, both for labour productivity and TFP, and across multiple industry splits. The main exception emerges when classifying leader firms by productivity, a more volatile measure. Under this alternative definition, the most productive firms do not slow down, but these firms are not particularly large in terms of employment and revenue, which limits their contribution to the aggregate slowdown (Section 3.4).

Turning to R&D patterns, we provide evidence that the productivity payoff to R&D spending has weakened over time. We first show that R&D intensity broadly rose across firms and industries, even as productivity growth declined. Because R&D is the primary channel through which firms invest in future productivity, this disconnect motivates a direct test of whether the translation of R&D expenditures into productivity gains has weakened. Estimating an R&D production function at both the industry and firm levels, we find that the elasticity of productivity with respect to R&D is consistently larger in the pre-period (1987–2006) than in the full sample (1987–2022), a pattern that holds across both productivity measures (TFP, labour productivity), different specifications, and multiple weighting schemes. These patterns are consistent with the "ideas getting harder to find" hypothesis and suggest that the slowdown may stem from shifts in the nature of the R&D process rather than changes in innovation effort.

2. Data and Measurement

This article begins by documenting aggregate trends and then decomposes them using more detailed firm- and industry-level data. We decompose aggregate manufacturing productivity into contributions across two groups of firms within each industry, which we label leader and follower firms, and across two groups of industries, which we refer to as frontier and laggard industries.

We consider three definitions for leader firms: employment (the baseline, capturing a fairly persistent proxy for size of inputs), revenue (highly correlated with employment but proxying size of output), and productivity (similar to the definition considered by Andrews *et al.* (2019) in the 'best-versus-rest' narrative). We use the average of the previous and current period (period-average) weights as the baseline and find similar results using either current-period and previous-period weights. We study the top-four frontier industries defined by labour productivity growth in the pre-period, 1987–2007, which are Textile mills (NAICS 313), Computer and electronic product manufacturing (NAICS 334), Electrical equipment, appliance and component manufacturing (NAICS 335) and Transportation equipment manufacturing (NAICS 336). In the Appendix, we provide additional details for NAICS 334 as the sole frontier industry. We evaluate both labour productivity and

TFP throughout the analysis. The reason for these different splits is to understand how pervasive the slowdown is given different productivity definitions and different cuts of the data.

2.1 Data Sources

Our analysis bridges industry-level and firm-level data for U.S. manufacturing (NAICS 31–33) over the 1987–2022 period. At the industry level, we draw on the BLS detailed industry productivity accounts (Bureau of Labor Statistics, 2025a,b) at NAICS 4-digit level, which provide labour productivity and TFP indices (both 2017=100), hours worked, sectoral output, and output price deflators. The BLS indices incorporate detailed-industry deflators, including hedonic adjustments for NAICS 334. At the firm level, we use Compustat North America (Standard & Poor's, 2025), drawing on revenue, employees, PP&E (Property, Plant, & Equipment), R&D (Research & Development), and cost of goods sold, restricting the sample to firms with non-missing employment and revenue and dropping bare 2-digit NAICS codes. For PP&E, we use historical prices consistent with financial statements and the existing R&D data come from U.S. Bureau of Economic Analysis (BEA) fixed-asset investment by industry (U.S. Bureau of Economic Analysis, 2025) and from Compustat at the firm level.

We deflate Compustat revenues by BLS NAICS 3-digit gross-output deflators, capital by major industry capital deflators, and R&D by college-educated male wages (Bloom *et al.*, 2020; U.S. Census Bureau, 2025). Details are in Appendix A.

To bridge the two data sources, we build explicit aggregations from the firm to the industry level and from the industry level to the level of the entire manufacturing sector. The firm-to-industry aggregation uses employment-share weights for labour productivity and revenue-share weights for TFP; industry-to-aggregate aggregation uses BLS hours shares for labour productivity and nominal output shares for TFP (see Appendix B.1 for details on the log-additive aggregation procedure). We construct aggregate manufacturing TFP as the weighted average of NAICS-4 digit industry TFP growth rates, with each industry weighted by its period-average share of total nominal manufacturing gross output. This procedure more closely aligns with how BLS calculates (sectoral output) TFP, and closely tracks the aggregate in Figure 1.[1] As we show below, the BLS and Compustat measures are closely aligned, which enables more granular analysis: the firm-level data can be used to decompose the industry-level trends without introducing measurement discrepancies.

2.2 Time Periods, Industries, and Firm Groups

This analysis primarily focuses on two time periods: the high-growth period (the "pre-period") in manufacturing (1987–2007) and the low-growth period (2010–2022). We exclude the Great

1 More details on the output measures are discussed in Eldridge and Powers (2023), who note the advantages of using sectoral output as the proper TFP measure.

Recession years 2008–2009. We thus reference 1987 as the initial date for the analysis. We then partition the data along two dimensions: an industry dimension that separates high-growth frontier industries from the rest of manufacturing, and a firm dimension that separates leaders from followers within each industry.

The firm and industry partitions serve as the main cuts of our analysis. The baseline industry split defines the top four NAICS 3-digit sectors by pre-period labour productivity growth as frontier industries, with the rest of manufacturing as laggard industries; see Section 3.2. As a robustness check, we also consider a narrower split that isolates Computer and electronic product manufacturing, which is by far the fastest-growing industry over 1987–2007 (Atalay *et al.*, 2025), as the sole frontier industry (Appendix C.2).

As for the two firm types, within each 4-digit NAICS code-year cell, we classify the top 10 per cent of firms as leaders (or the single largest firm if fewer than 10 are present); the remainder are followers. We consider three ranking variables. Employment (the baseline) captures scale and is persistent over time; revenue captures market position and is highly correlated with employment. Both are more persistent than productivity rankings (Pearce and Wu, 2025). Productivity (labour productivity or TFP) captures the technological frontier but is much less persistent: smaller firms are often the most productive. For defining leaders, we base our measure on an average of the current and previous period.

2.3 Productivity and R&D

We measure labour productivity at the firm level as deflated revenue per employee. At the industry level, we use the BLS labour productivity index (2017=100) and the BLS Total Factor Productivity index. These measures are relevant for our industry analysis.

For firm-level TFP, we follow the same approach used by De Ridder *et al.* (2026), who estimate a gross-output production function with two inputs: a variable input (cost of goods sold, absorbing both materials and labour payments) and a fixed input (net PP&E). Specifically, we define:

$$\ln TFP_{ft} = \ln Y_{ft} - \hat{\beta}_{X,j} \ln X_{ft} - \hat{\beta}_{K,j} \ln K_{ft}, \quad (1)$$

where Y_{ft} is deflated revenue, X_{ft} is deflated cost of goods sold, K_{ft} is deflated net PP&E, and the elasticities $(\hat{\beta}_{X,j}, \hat{\beta}_{K,j})$ are estimated at the NAICS 3-digit industry codes following the De Ridder *et al.* (2026) procedure. As for the industry-level productivity measures, we directly use the BLS TFP Index rather than aggregating firm-level estimates, so that industry-level productivity in the R&D regressions (Section 4) is measured independently of the Compustat data.

We measure R&D intensity as real R&D per worker, or R&D relative to revenue. These measures are discussed in greater detail in Section 4. For R&D expenditures, the BEA only provides details at the NAICS 3-digit level as far as we are aware. In order to get more granular industry analysis, we extend to NAICS 4-digit level using Compustat data.

3. Productivity Trends across Firms and Industries

This section documents the aggregate manufacturing productivity slowdown and decomposes its drivers at the industry and firm levels. Using the four-group framework introduced in Section 2, we examine whether the slowdown is concentrated in particular groups or spread across the distribution. The central finding is that the slowdown is pervasive: under size-based rankings (employment or revenue), both leaders and followers slow down. Section 3.4 examines the alternative productivity-based ranking, which yields a different pattern.

3.1 Aggregate Slowdown

We begin by documenting the aggregate slowdown. Figure 1 plots labour productivity and TFP indices for both BLS and Compustat, each initialized to 100 in 1987. For our BLS series (Figures 1A and 1C), we plot our industry-level aggregation, with current shares and fixed shares as well as the two BLS aggregates for labour productivity and TFP. We find that our aggregation shows a very similar trend to the BLS aggregate manufacturing method.

The picture is striking: labour productivity grew at an annualized rate of roughly 3.5 per cent per year over the 1987–2007 period, approximately doubling in level by 2007, and then stagnated, with annualized growth falling to near zero over the 2010–2022 period. TFP grew at roughly 1.3 per cent per year through 2007 and then flattened. The slowdown is visible in every panel of Figure 1: all productivity indices rise steeply through the pre-period, and the change in slope around 2007 marks the onset of the slowdown across both productivity measures and both data sources. The two data sources are closely aligned under both measures, confirming that the Compustat sample of publicly traded firms captures the same broad trends as the BLS universe.

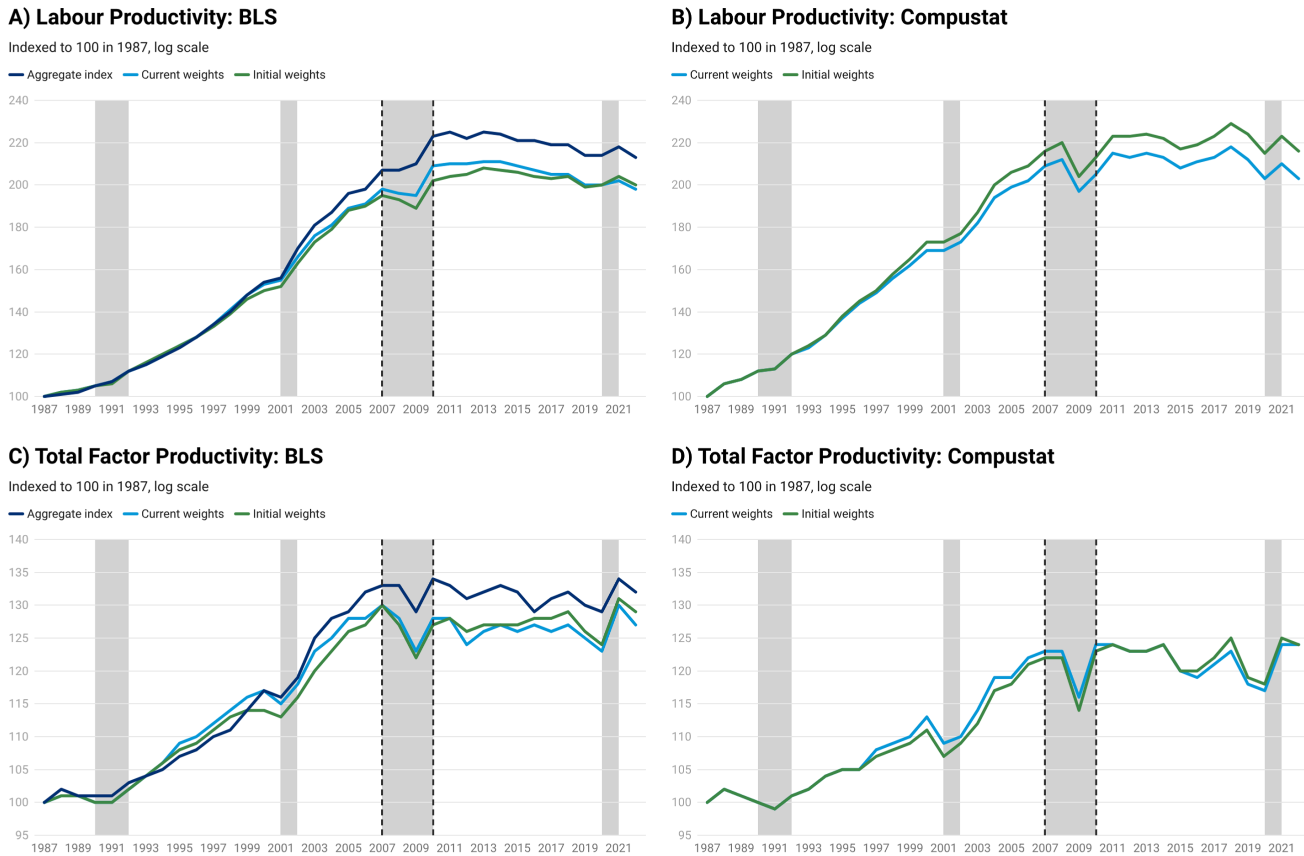
Table 1 quantifies the slowdown. Annualized labour productivity growth fell by roughly 4 percentage points from the pre-period (1987–2007) to the post-period (2010–2022), whether measured from BLS or Compustat. TFP slows by 1.1–1.4 percentage points, consistent with the labour productivity pattern but more muted. Both BLS and Compustat TFP are near zero or slightly negative in the post-period. Which firms and industries account for the flattening?

3.2 Within-Group Trends

We start by focusing on industry- and firm-level splits in accumulated growth from 1987–2022. This section focuses on overall trends and average growth within each group before turning to each group's contribution to aggregate productivity growth, which depends on its size.

Figures 2 and 3 plot each group's own labour productivity and TFP trajectories under period average weights for employment- and revenue-ranked leaders, using the top four frontier industries versus the rest. The common trend is immediately apparent: all four groups (leader and follower firms in both the frontier and laggard industry) rise together through the pre-period and flatten together after 2007.

Figure 1: Aggregate Manufacturing Productivity Indices



Note: Gray shaded areas denote NBER recession periods. Vertical dashed lines indicate the subperiod break points.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

Table 1: Aggregate Manufacturing Productivity Slowdown

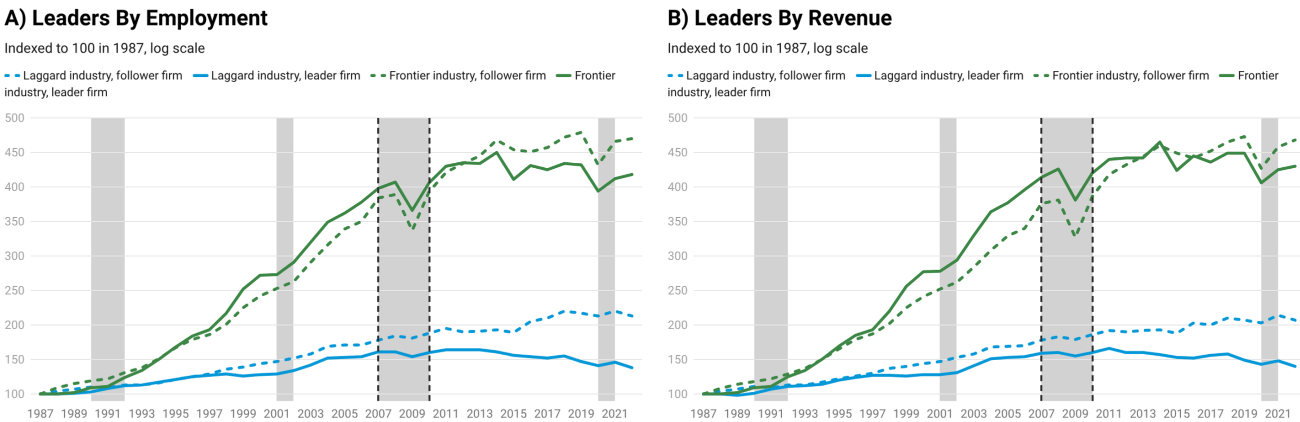
Log growth rates, average annual per cent change

Annualized Growth	1987-2007	2010-2022	Slowdown
Labour productivity (BLS)	3.46	−0.42	−3.88
Labour productivity (Compustat)	3.98	−0.07	−4.05
Total factor productivity (BLS)	1.33	−0.05	−1.38
Total factor productivity (Compustat)	1.05	−0.04	−1.09

Note: Slowdown = Post − Pre period change in growth rates.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

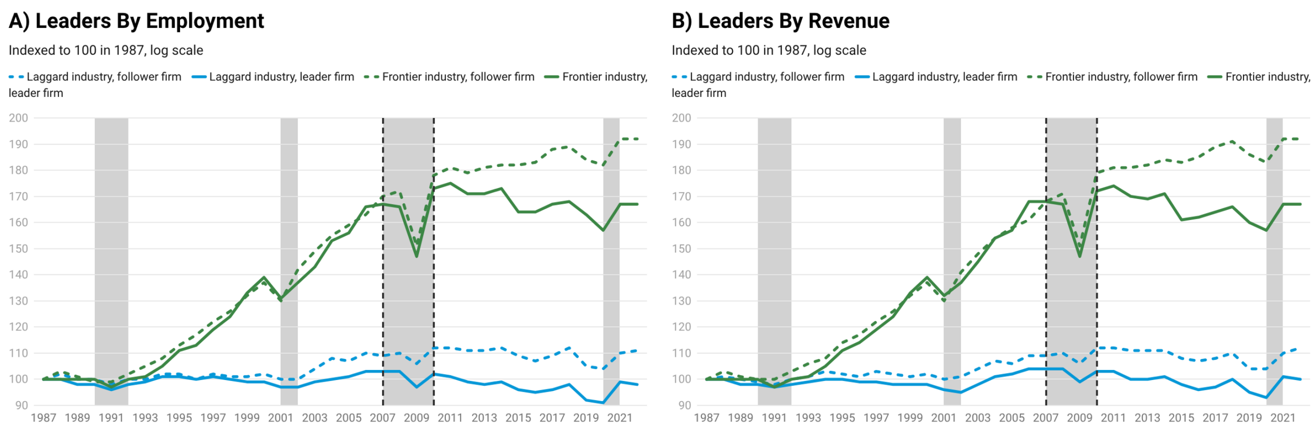
Figure 2: Group-Level Labor Productivity Trajectories (Top 4 Versus Rest)



Note: Each panel plots the own labor productivity index for each of the four firm × industry groups (top four NAICS 3-digit sectors by pre-period labor productivity growth (313, 334, 335, 336) vs. rest, leader vs. follower firm). Shares and leader classifications are based on period-average weights, $\bar{\omega}_{it} = (\omega_{it} + \omega_{i,t-1})/2$. Leaders = top 10% within NAICS 4-digit × year by employment (left) or revenue (right). Gray shaded areas denote NBER recession periods.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

Figure 3: Group-Level Total Factor Productivity Trajectories (Top 4 Versus Rest)



Note: Same as Figure 2 but for Total Factor Productivity.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

No single group pulls away or collapses on its own. The narrower 334-only industry split yields qualitatively similar results (Appendix C.2). Section 3.4 examines what happens when leaders are instead defined by productivity. The qualitative results are the same under different weightings and leading firm definitions (see Appendix C.1 and Figure 12 for additional robustness checks).

Table 2 reports within-group growth rates for the top four versus rest industry split. Under employment rankings (Table 2), all four groups slow down. The frontier industries (both leaders and followers) grew rapidly in the pre-period and then sharply decelerated. Leaders and followers in laggard industries also slowed. For TFP, the pattern is qualitatively similar: frontier firms slowed substantially while laggard-industry firms slowed more modestly. The slowdown is present in both leaders and followers in laggard industries.

Table 2: Growth by Group: Industry and Firm Groups

Log growth rates, average annual per cent change

		Panel A: Labour Productivity	
Annualized Growth	**1987-2007**	**2010-2022**	**Slowdown**
Frontier industry: leader firms	7.15	0.24	-6.91
Frontier industry: follower firms	6.97	1.49	-5.48
Lagged industry: leader firms	2.39	-1.21	-3.60
Laggard industry: follower firms	2.94	1.04	-1.90
		Panel B: Total Factor Productivity	
Annualized Growth	**1987-2007**	**2010-2022**	**Slowdown**
Frontier industry: leader firms	2.58	-0.31	-2.90
Frontier industry: follower firms	2.68	0.63	-2.06
Lagged industry: leader firms	0.16	-0.28	-0.45
Laggard industry: follower firms	0.45	-0.08	-0.54

Note: Annualized log growth rates (average annual per cent change) by group, period-average weights. Slowdown = Post – Pre. Leaders = top 10 per cent by employment within NAICS 4-digit × year. Frontier = top four NAICS 3-digit sectors by pre-period labour productivity growth (313, 334, 335, 336).

Source: Authors' calculations using Compustat and BLS data. • Created with Datawrapper

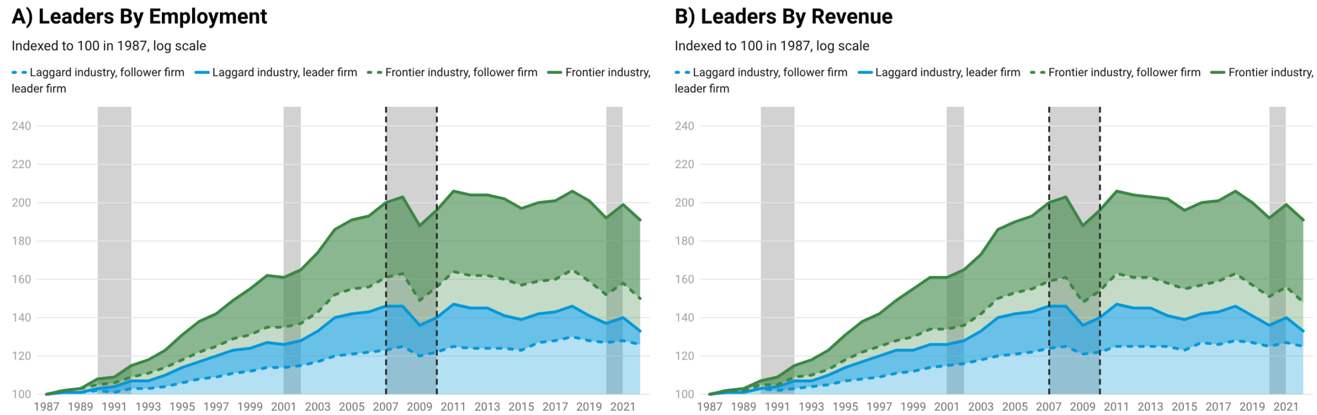
3.3 Decomposing the Aggregate Slowdown

The trajectory figures show that all groups slow down in productivity, but these groups differ in their contribution to aggregate growth in manufacturing. To assess how much each group contributes to the aggregate slowdown, we next turn to decompositions of aggregate productivity to each of these groups.

Figure 4 presents the central result for the top four industries versus rest split. Each panel shows the decomposition of aggregate labour productivity. The visual message mirrors the trajectories: all four bands rise together and flatten together. Both leaders and followers contribute to the rise and to the subsequent stagnation. Figure 5 repeats the exercise for TFP; the same pattern holds throughout. The narrower 334-only split yields qualitatively similar results (Appendix C.2). These patterns persist across different weighting methods (see Appendix C.1, Figures 9–10 for current-period, previous-period, and period-average weights side by side).

We decompose aggregate productivity

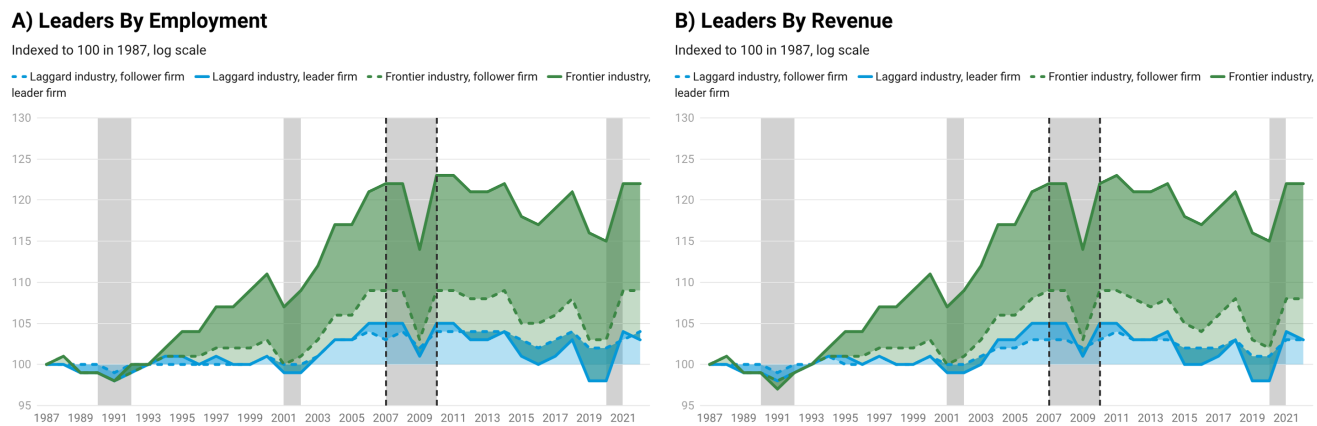
Figure 4: Four-Group Labour Productivity Decomposition (Top 4 Versus Rest)



Note: Stacked-area decomposition (log method, period-average weights) of aggregate labor productivity into four groups (leader/follower firm × frontier/laggard industry). Frontier = top four NAICS 3-digit sectors by pre-period labor productivity growth (313, 334, 335, 336). Leaders = top 10% within NAICS 4-digit × year by employment (left) or revenue (right). Gray shaded areas denote NBER recession periods.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

Figure 5: Four-Group Total Factor Productivity Decomposition (Top 4 Versus Rest)



Note: Same as Figure 4 but decomposing aggregate TFP rather than labor productivity. Gray shaded areas denote NBER recession periods.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

growth using the log-additive approximation, which enables us to aggregate the separate contribution of each group ignoring the reallocation term. The method is consistent with the fact that aggregate log productivity growth is close to a weighted sum of group-level log productivity growth, $\Delta \log LP_t \approx \sum_g \omega_{gt}, \Delta \log LP_{gt}$, where ω_{gt} is group g's share of aggregate hours. Each group's contribution to aggregate growth is then its hours share times its own productivity growth. Slowdown shares, defined as each group's share of the preto-post change in aggregate growth, sum to one by construction. This is only an approximation, but the small differences between initial-share productivity growth and evolving shares in Table 1 indicate it is unlikely to generate strong divergence. As a robustness check, we also report a level-additive (exact) decomposition in Appendix B.1. The qualitative conclusions are the same under both methods.

These results are robust to the choice of weighting scheme.[2] The qualitative results are the same under all three schemes (see Appendix C.1, Figures 9–10). To assess the extensive margin directly, Appendix C.1 reports a Foster *et al.* (2001) decomposition that explicitly separates the contributions of entering, exiting, and continuing firms.

All groups contribute to the slowdown. Leader firms in the frontier industries account for about a quarter of the labour productivity slowdown and nearly half the TFP slowdown, but leaders and followers in laggard industries together account for roughly 70 per cent of the labour productivity slowdown (see Table 10 in Appendix C.2 for the full breakdown). We next focus on a split that has been discussed in the literature, that of looking at the most productive firms rather than the largest in a given industry.

3.4 Comparing to Andrews *et al.* (2019)

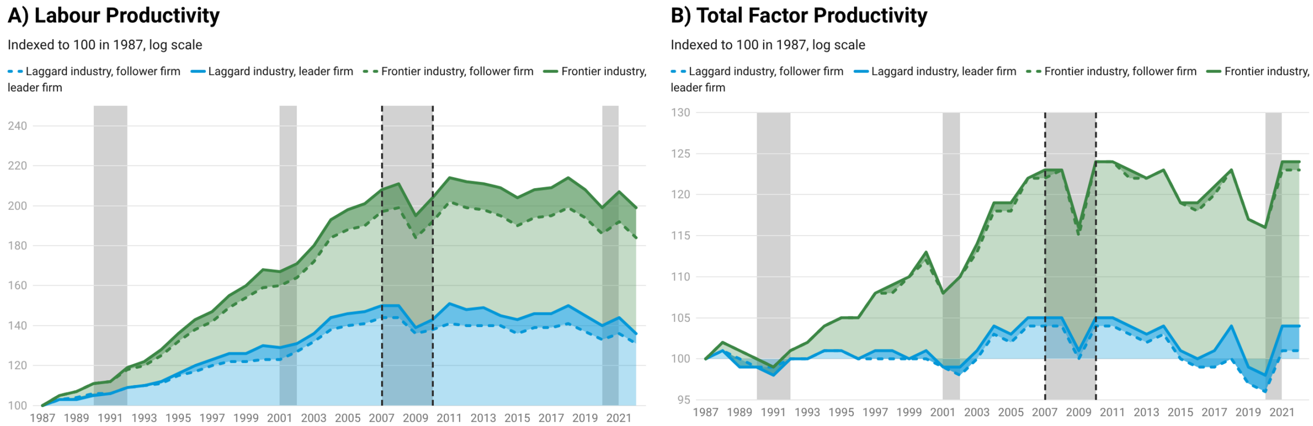
The preceding analysis defined leaders by employment or revenue, both of which are persistent measures of firm size. An alternative, proposed by Andrews *et al.* (2019), defines the frontier by productivity itself. Under this definition, the pattern changes: the most productive firms continue to grow, while followers stagnate. However, productivity-ranked leaders are a different set of firms than size-ranked leaders: they are smaller, less persistent from one period to the next, and command a much smaller share of aggregate employment and revenue. Their continued growth therefore contributes little to any aggregate dynamics.

Figure 6 presents the four-group decomposition under productivity ranking for labour productivity and TFP. The follower bands dominate: they account for nearly all of the rise and the subsequent stagnation, while the leader bands remain thin. The most productive firms in the top four industries actually offset the slowdown, contributing positively after 2007. This contrasts with Figures 4–5, where leading firms also experienced significant slowdowns.

Table 3 quantifies the contrast. Productivity-ranked leaders in the frontier industries still grew in the post-period, though at a substantially lower rate than in the pre-period. Productivity-ranked leaders in laggard industries experienced a relatively modest slowdown. Meanwhile, productivity ranked followers in laggard industries actually declined in the post-period. For TFP, the shift is even more striking: TFP-ranked leaders in laggard industries actually accelerated, offsetting the slowdown elsewhere. Follower firms account for 88 per cent of the labour productivity slowdown under this ranking (Table 11 in Appendix C.2).

2 Our method restricts the sample to continuing firms, since entrants and exiters lack the lagged or current shares needed to form the average. Our restriction implicitly attributes to entrants and exiters the same productivity growth as the continuing-firm average. If entry and exit patterns shifted between the pre- and post-periods, the period-average decomposition would miss this channel. Under current-period weights, entrants appear with their current share but contribute zero within-firm growth; the between term then picks up some entry/exit composition effects but conflates them with reallocation among incumbents. Whereas in Lashkari and Pearce (2026) we use current-period weights as the baseline, here we switch to period-average weights. See Appendix C.1 for further discussion.

Figure 6: Four-Group Decomposition Under Productivity Ranking (Top 4 Versus Rest)



Note: Stacked-area decomposition (log method, period-average weights) of aggregate productivity into four groups (leader/follower firm × frontier/laggard industry). Leaders = top 10% within NAICS 4-digit × year by own labor productivity (left) or TFP (right). Frontier = top four NAICS 3-digit sectors by pre-period labor productivity growth (313, 334, 335, 336). Gray shaded areas denote NBER recession periods.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

Table 3: Within-Group Growth: Productivity-Ranked Leaders

Log growth rates, average annual per cent change

Panel A: Labour Productivity

Annualized Growth	1987-2007	2010-2022	Slowdown
Frontier industry: leader firms	8.68	4.28	-4.40
Frontier industry: follower firms	7.16	0.65	-6.51
Lagged industry: leader firms	5.74	2.13	-3.62
Laggard industry: follower firms	2.76	-0.76	-3.52

Panel B: TFP

Annualized Growth	1987-2007	2010-2022	Slowdown
Frontier industry: leader firms	2.99	0.47	-2.52
Frontier industry: follower firms	2.77	0.21	-2.56
Lagged industry: leader firms	-1.40	3.18	4.58
Laggard industry: follower firms	0.29	-0.24	-0.54

Note: Same as Table 2 but leaders defined as top 10 per cent by own productivity (labour productivity or TFP) within NAICS 4-digit × year.

Source: Authors' calculations using Compustat and BLS data. • Created with Datawrapper

The key reason productivity-ranked leaders matter less for the aggregate is mechanical: they command a small share of aggregate employment, so their continued growth has little weight in any decomposition. This finding is inconsistent with a "frontier exhaustion" story in which the most productive firms run out of ideas, and more consistent with broad-based mechanisms affecting the bulk of the firm distribution.

4. R&D and Research Productivity

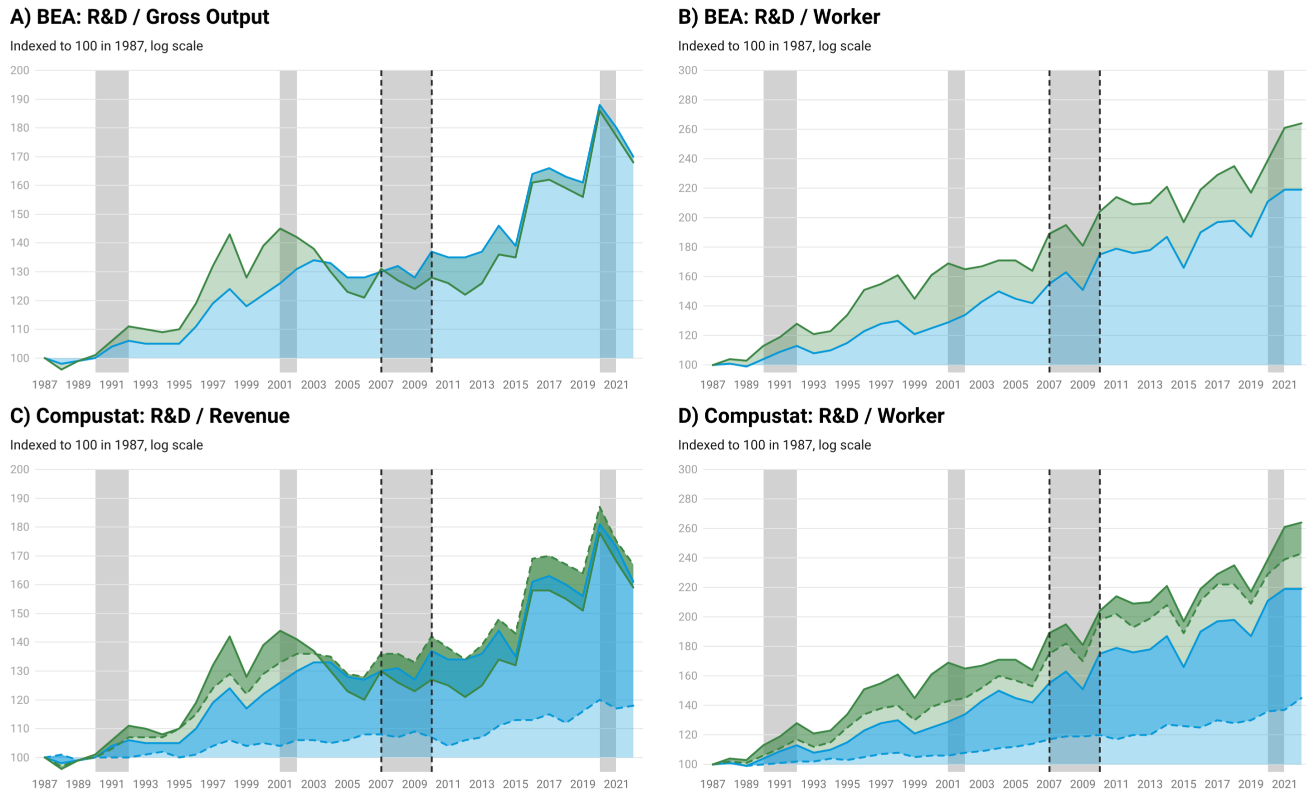
The preceding section established that the productivity slowdown is generally broad-based. In Lashkari and Pearce (2026), we documented that R&D intensity rose even as productivity growth stalled, but we did not examine whether there has been a change in the innovation production function that maps R&D expenditures into productivity growth. In this section, we expand our analysis of the trends in R&D intensity and further study the relationship between R&D spending and productivity growth.

4.1 Trends in R&D Intensity

Figure 7 decomposes R&D intensity into contributions from frontier industries (top 4) and laggard industries, using both industry-level (BEA/BLS) and firm-level (Compustat) data. The top row measures R&D intensity at the industry level, dividing the BEA fixed-asset measure of R&D investment by BLS gross output (panel a) and BLS employment (panel b), with shaded areas distinguishing the contribution of frontier (green) and laggard (blue) industries. The bottom row applies the same decomposition at the firm level using Compustat data, further splitting each industry group into leader and follower firms within each NAICS 4-digit industry.

Several patterns emerge from Figure 7. First, R&D intensity rose broadly: the intensity in every group in every panel is higher in 2022 than in 1987, confirming that the productivity slowdown cannot be attributed to a decline in relative innovative effort. Second, the industry-level panels reveal that frontier industries consistently maintained higher R&D intensity than laggard industries, with the gap widening after 2000, particularly in R&D per worker (panel b). Third, the firm-level decomposition adds a within-industry dimension: in laggard industries, follower firms (light blue) account for most of the R&D intensity growth, while leader firms within those industries grew more modestly. In frontier industries, the pattern is reversed, with leader firms (dark green) driving the increase. Fourth, the R&D-to-gross-output ratio in panel (a) shows a pronounced dip around 2010 in frontier industries, mirroring the dot-com bust pattern from 2001, but recovers sharply afterward. Together these panels establish that the post-2007 productivity slowdown coincided with continued, and in most measures accelerating, growth in R&D effort across all groups. The natural question is therefore not whether firms stopped investing in innovation, but whether those investments became less effective at generating productivity gains.

Figure 7: R&D Intensity Decomposition: Frontier (Top 4) Versus Laggard Industries



Note: All series indexed to 1987 = 100; log scale. Top row: industry-level R&D intensity from BEA fixed-asset investment divided by BLS gross output (a) and BLS employment (b), decomposed into frontier industries (top 4, blue) and laggard industries (red). Bottom row: firm-level R&D intensity from Compustat, measured as R&D expenditure divided by revenue (c) and employment (d), decomposed into four groups by industry (frontier vs. laggard) and firm rank (leader vs. follower within NAICS 4-digit). Gray bars indicate NBER recession dates; dashed vertical lines mark 2007 and 2010. R&D intensity rises in all groups throughout the sample.

Source: Authors' calculations using BLS and Compustat data. • Created with Datawrapper

4.2 Estimating the R&D–Productivity Connection

The preceding sections document two facts: the productivity slowdown is broad-based, and R&D intensity has risen even as productivity growth stalled. The natural question is whether these two facts are connected: has R&D become less effective at generating productivity growth? To answer this question, we need a framework that links R&D spending to subsequent productivity growth, controlling for the persistence of productivity itself. The dynamic panel literature provides such a framework (Arellano and Bond, 1991; Blundell and Bond, 1998).

We model the relationship as an R&D production function in which current R&D impacts the evolution of productivity in the next period following:

$$\ln Z_{i,t+1} = \alpha \ \ln R_{it} + \phi \ \ln Z_{it} + \eta_i + u_{it} \tag{2}$$

where t indexes consecutive 4-year periods beginning with 1987–1990, Z_{it} is the mean productivity of unit i over period t (labour productivity LP or total factor productivity TFP), R_{it} is the mean real R&D expenditure over the same period (deflated by skilled wages following Bloom *et al.* 2020), and η_i is a unit fixed effect. The coefficient α captures the elasticity of next-period productivity with respect to current R&D, controlling for the persistence of productivity (ϕ). A decline in α from the pre-period to the full sample would indicate that a given increase in R&D spending translates into less productivity growth than it did before, consistent with the "ideas getting harder to find" hypothesis (Bloom *et al.*, 2020).[3]

Estimating (2) by OLS with unit fixed effects would introduce Nickell bias: the within-group transformation mechanically correlates the lagged dependent variable with the error, biasing ϕ downward and potentially contaminating α (we report OLS estimates in Appendix D for comparison). We address this issue using the system GMM estimator of Arellano and Bond (1991) and Blundell and Bond (1998). The idea is to estimate the model in two forms simultaneously. In first differences, the unit fixed effect η_i drops out, and lagged levels of $\ln Z_{it}$ and $\ln R_{it}$ serve as instruments for the differenced regressors. In levels, lagged differences of $\ln Z_{it}$ and $\ln R_{it}$ serve as instruments, exploiting the assumption that these differences are uncorrelated with the fixed effect. Combining both sets of moment conditions yields more efficient estimates than either alone, particularly when the series are persistent, as productivity tends to be.

We estimate (2) at both the firm and the industry level. We note, however, that the industry level estimation has the advantage of capturing not only within-firm returns to R&D but also spillovers and business-stealing effects that operate across firms within an industry; a decline in α at the industry level therefore reflects a weakening of the overall effect of R&D on industry productivity, inclusive of these channels.

At the industry level, the unit of observation is a 4-digit NAICS industry code, productivity is measured directly from BLS indices, and R&D expenditure is aggregated from Compustat to the 4-digit NAICS level. We consider three BLS productivity measures: TFP, real output per hour (the BLS's preferred labour productivity measure), and real output per worker (more directly comparable to the firm-level measure). At the firm level, the unit of observation is a Compustat firm, and we estimate the model for both TFP, computed as a Solow residual following De Ridder*et al.* (2026), and labour productivity (deflated revenue per worker).

We estimate each specification under several variations to ensure robustness. We report unweighted and weighted estimates (using hours, employment, or nominal output weights at the industry level, and employment or revenue weights at the firm level), with and without time-period fixed effects, and under two instrument sets: all available lags (two or more) and only the first two available lags (two to three), the latter guarding against instrument proliferation (Roodman, 2009). For each specification, we report the Arellano and Bond (1991) tests for first- and second-order

3 The magnitude of our industry-level elasticities is broadly consistent with Doraszelski and Jaumandreu (2013), who estimate firm-level R&D elasticities below 0.05 using Spanish manufacturing data. Other studies report somewhat different magnitudes depending on the time period, country, and whether they focus on within-firm returns or broader spillover channels (Peters *et al.*, 2013; Fieldhouse and Mertens, 2023; Dyevre, 2024).

serial correlation in the differenced residuals (significant AR(1) is expected by construction, while significant AR(2) would signal misspecification) and the Sargan and Hansen tests for joint instrument validity.

We estimate the model over the full sample (1987–2022) and then re-estimate it for the pre-period alone (1987–2006, to match the 4-year period definitions). Comparing estimates of α across the two samples is the central exercise: if R&D has become less effective, the pre-period estimate should exceed the full-sample estimate. We do not estimate the model separately for the post-period because we are left with a short panel in this period (only three non-overlapping 4-year periods after excluding 2008–2009), leaving the GMM moment conditions underpowered and leading to generally implausible parameter estimates with wide standard errors.

4.3 Estimation Results

Table 4 reports GMM estimates of the specification in Equation (2) at the industry level, comparing the pre-period to the full sample for total factor productivity. We report the specification for labour productivity in Appendix D. The results are similar across both specifications.

The R&D coefficient $\hat{\alpha}$ is larger in the pre-period in all specifications, declining by roughly one-third to one-half when the sample is extended to include the post-period. Results using output per worker are reported in Appendix D and are similar.

The diagnostic tests support the GMM specification. The AR(2) test, which is the key check on whether the moment conditions are valid, fails to reject the null hypothesis in all specifications for both labour productivity and TFP, indicating no evidence of second-order serial correlation in the differenced residuals. The Sargan test rejects the null hypothesis throughout, but this is expected: the Sargan statistic assumes homoskedastic errors and is known to over-reject in the presence of heteroskedasticity (Roodman, 2009). The Hansen test, which is robust to heteroskedasticity, passes in specifications that include time fixed effects, but rejects in some specifications without time fixed effects. Since the point estimates of α are stable across specifications with and without time fixed effects, and the AR(2) test uniformly passes, we interpret the Hansen rejections as reflecting residual heteroskedasticity rather than fundamental instrument invalidity.

We present the full set of firm-level results in Appendix D. The firm-level results also exhibit patterns similar to the industry-level results, albeit somewhat less uniformly. The decline in $\hat{\alpha}$ is clear in specifications without time fixed effects (for instance, the unweighted labour productivity baseline drops from 0.103 to 0.066). When time fixed effects are included, they absorb common temporal variation and compress $\hat{\alpha}$ toward zero in both periods, so the pre-versus-full gap narrows. The firm estimates may be less reliable, because spillovers from R&D expenditures are significant and thus less responsive to firm-level decisions than industry-level changes.

We also estimate Equation (2) using OLS with unit and time-period fixed effects (see Appendix D). The OLS point estimates of α are smaller than the GMM

Table 4: R&D Production Function: Industry-Level Total Factor Productivity, Pre vs. Full

	Pre	Full	Pre	Full	Pre	Full	Pre	Full
$\hat{\phi}$	0.728	0.705	0.717	0.679	0.730	0.714	0.737	0.714
	(0.046)	(0.035)	(0.041)	(0.040)	(0.034)	(0.034)	(0.047)	(0.036)
$\hat{\alpha}$	0.0183	0.0132	0.0267	0.0131	0.0144	0.0075	0.0258	0.0122
	(0.0096)	(0.0059)	(0.0134)	(0.0092)	(0.0056)	(0.0053)	(0.0107)	(0.0069)
Observations	318	619	318	619	318	619	318	619
Weighted			X		X		X	X
Time fixed effects					X	X	X	X
AR (1)	0.024	0	0.477	0.033	0.303	0.000	0.967	0.054
AR (2)	0.410	0.644	0.851	0.313	0.310	0.223	0.265	0.099
Sargan	0	0	0	0	0	0	0	0
Hansen	0	0.004	0	0.014	0.179	0.026	0.203	0.230

Note: Same as Table 12 but using TFP (BLS index) as the productivity measure. Weighted specifications use nominal output weights.
Source: Authors' calculations using Compustat and BLS data. • Created with Datawrapper

estimates, as expected given Nickell bias, but the pre-period estimates still exceed the full-sample estimates at the industry level across all specifications.

The decline in $\hat{\alpha}$ from the pre-period to the full sample is the central finding of this section. At the industry level, the pattern is pervasive: it holds across all three productivity measures, all weighting and instrument choices, with and without time fixed effects, and under both GMM and OLS. At the firm level, the decline is present in specifications without time fixed effects, but it is attenuated when time fixed effects absorb the common slowdown. Whether the decline reflects a structural shift in the knowledge production function or a compositional change (for instance, R&D shifting toward harder problems or toward activities that are not well captured by standard productivity measures) is a question these reduced-form estimates cannot resolve. What they do establish is that the relationship between R&D spending and subsequent productivity growth has weakened. Combined with the rising R&D intensity documented in Figure 7, this suggests that the manufacturing productivity slowdown reflects declining research productivity rather than reduced innovation effort.[4]

5. Conclusion

This article asks two questions about

4 Alternatively, Ando *et al.* (2025) find that ideas are not getting harder to find. Higher rates of knowledge obsolescence or lower spillovers would be consistent with this story and declining productivity growth.

the U.S. manufacturing productivity slowdown. First, does it emerge broadly among large and small firms and among frontier and laggard industries? The answer is yes; the slowdown is evident across both groups of industries and across both leader and follower firms. This finding is robust across different weighting methods, different productivity measures (labour productivity and TFP), and multiple definitions of frontier/laggard industries or leader/follower firms. Firms selected in each period as most productive in their respective industries are an exception in that they do not exhibit a productivity slowdown, however, they only account for a small share of aggregate employment and revenue, limiting their role in any aggregate decomposition.

Second, does the slowdown reflect declining R&D effectiveness? R&D intensity rose across firms and industries even as productivity growth declined. System GMM estimates of an R&D production function at both the industry and firm levels show that the elasticity of productivity growth with respect to R&D expenditures is consistently weaker in the full sample than in the pre-period. These findings are consistent with the "ideas getting harder to find" hypothesis and suggest that the problem is not reduced innovation effort, but reflects a weakening link between R&D spending and productivity growth.

The key contrast in our findings is between two narratives. Leader-divergence theories (Andrews *et al.*, 2019; Aghion *et al.*, 2023) predict that the slowdown originates in follower firms, who fall further behind advancing technological leaders. Defining the frontier by productivity indeed yields a decomposition consistent with this story, but productivity-ranked leader firms are small and account for a negligible share of aggregate employment and revenue, implying that the divergence narrative does not explain the aggregate productivity slowdown. Similarly, rising-concentration accounts (Olmstead-Rumsey, 2022; Klenow *et al.*, 2019) predict that the slowdown should be concentrated among smaller firms losing ground to dominant ones; our decomposition shows that both large and small firms slow down under size-based rankings. The broad-based pattern is instead more consistent with mechanisms that affect the entire firm distribution, as in the declining-dynamism framework of Akcigit and Ates (2023) or the "ideas getting harder to find" view of Bloom *et al.* (2020). Our R&D production function estimates provide direct evidence for the latter channel: the problem is not that firms stopped investing in research, but that a given dollar of R&D buys less productivity growth than it once did.

Two questions follow naturally. First, can manufacturing productivity bounce back? Our findings suggest that the declining effectiveness of R&D is a structural feature of the past two decades, however, the arrival of general-purpose technologies, such as artificial intelligence, has the potential to revitalize the frontier. Second, might the same patterns eventually hold for services, or will service-sector productivity continue to grow? This article cannot answer these questions, but it does provide an angle, linking Compustat firm-level data to industry-level aggregates, that can be applied to other sectors and time periods as data become available.

References

Aghion, P., A. Bergeaud, T. Boppart, P.J. Klenow, and H. Li (2023) "A Theory of Falling Growth and Rising Rents," *Review of Economic Studies*, Vol. 90, No. 6, pp. 2675–2702.

Akcigit, U. and S.T. Ates (2023) "What Happened to US Business Dynamism?" *Journal of Political Economy*, Vol. 131, No. 8, pp. 2059–2124.

Ando, Y., J.E. Bessen, and X. Wang (2025) "The Rising Returns to R&D: Ideas are not getting harder to find," Boston Univ. School of Law Research Paper, No. 25-16.

Andrews, D., C. Criscuolo, and P. Gal (2015) "Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries," Productivity Working Papers 2, OECD.

Andrews, D., C. Criscuolo, and P.N. Gal (2019) "The Best versus the Rest: Divergence across Firms during the Global Productivity Slowdown," Technical Report, Centre for Economic Performance, LSE.

Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, Vol. 58, No. 2, pp. 277–297.

Atalay, E., A. Hortaçsu, N. Kimmel, and C. Syverson (2025) "Why Is Manufacturing Productivity Growth So Low?" Technical Report, National Bureau of Economic Research.

Bloom, N., C.I. Jones, J. Van Reenen, and M. Webb (2020) "Are Ideas Getting Harder to Find?" *American Economic Review*, Vol. 110, No. 4, pp. 1104–1144.

Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 87, No. 1, pp. 115–143.

Bureau of Labor Statistics (2025a) "Total Factor Productivity by Detailed Industries in Manufacturing, Air Transportation, and Line-Haul Railroads." Available at: https://www.bls.gov/productivity/tables/. Accessed December 10, 2025.

Bureau of Labor Statistics (2025b) "Total Factor Productivity by Major Industries." Available at: https://www.bls.gov/productivity/tables/. Accessed December 10, 2025.

Byrne, D.M., J.G. Fernald, and M.B. Reinsdorf (2016) "Does the United States Have a Productivity Slowdown or a Measurement Problem?" *Brookings Papers on Economic Activity*, pp. 109–182.

De Ridder, M. (2024) "Market Power and Innovation in the Intangible Economy," *American Economic Review*, Vol. 114, No. 1, pp. 199–251.

De Ridder, M., B. Grassi, and G. Morzenti (2026) "The Hitchhiker's Guide to Markup Estimation: Assessing Estimates from Financial Data," *Econometrica*, Vol. 94, No. 1, pp. 137–168.

Doraszelski, U. and J. Jaumandreu (2013) "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, Vol. 80, No. 4, pp. 1338–1383.

Dyevre, A. (2024) "Public R&D Spillovers and Productivity Growth," Working Paper, LSE Department of Economics.

Eldridge, L.P. and S.G. Powers (2023) "Industry contributions to productivity growth in US manufacturing," *Monthly Labor Review*.

Fernald, J.G. (2015) "Productivity and Potential Output before, during, and after the Great Recession," *NBER Macroeconomics Annual*, Vol. 29, No. 1, pp. 1–51.

Fieldhouse, A.J. and K. Mertens (2023) "The Returns to Government R&D: Evidence from US Appropriations Shocks," Working Paper 2305, Federal Reserve Bank of Dallas.

Foster, L., J.C. Haltiwanger, and C.J. Krizan (2001) "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in *New Developments in Productivity Analysis*, University of Chicago Press, pp. 303–372.

Klenow, P.J., H. Li, and T. Naff (2019) "Is Rising Concentration Hampering Productivity Growth?" *FRBSF Economic Letter*, Vol. 28.

Lashkari, D. and J. Pearce (2024) "The Mysterious Slowdown in US Manufacturing Productivity," Liberty Street Economics.

Lashkari, D. and J. Pearce (2026) "The Productivity Paradox in US Manufacturing: A Firm- and Industry-Level Perspective," *AEA Papers and Proceedings*, forthcoming.

Olmstead-Rumsey, J. (2022) "Market Concentration and the Productivity Slowdown," Working Paper.

Ottonello, P. and T. Winberry (2020) "Financial Heterogeneity and the Investment Channel of Monetary Policy," *Econometrica*, Vol. 88, No. 6, pp. 2473–2502.

Pearce, J. and L. Wu (2025) "Market Concentration and Aggregate Productivity: The Role of Demand," Staff Reports 1159, Federal Reserve Bank of New York.

Peters, B., M.J. Roberts, V.A. Vuong, and H. Fryges (2013) "Estimating Dynamic R&D Demand: An Analysis of Costs and Long-Run Benefits," Working Paper 19374, National Bureau of Economic Research.

Roodman, D. (2009) "How to Do xtabond2: An Introduction to Difference and System GMM in Stata," *Stata Journal*, Vol. 9, No. 1, pp. 86–136.

Sharpe, A. and R. Chittoor (2025) "The Stylized Facts of the US Manufacturing Productivity Slowdown," Working Paper.

Standard & Poor's (2025) "Compustat North America." Distributed by Wharton Research Data Services. Accessed December 10, 2025.

Syverson, C. (2017) "Challenges to Mismeasurement Explanations for the US Productivity Slowdown," *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 165–186.

U.S. Bureau of Economic Analysis (2025) "Detailed Data for Fixed Assets and Consumer Durable Goods: Nonresidential Detailed Investment — Investment." Available at: https://apps.bea.gov/iTable.

U.S. Census Bureau (2025) "Historical Income Tables: People, Table P-18: Educational Attainment — People 25 Years Old and Over by Mean Income and Sex." Available at: https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-people.html. Accessed December 10, 2025.

Productivity Growth in the U.S. Medical Care Sector: An Analysis Using the BEA's Health Care Satellite Account

Calvin Ackley, Abe Dunn, Eli Liebman and John A. Romley*

IPM Research Article

Abstract

Understanding health care productivity is critical, as the sector accounts for about 17 per cent of U.S. gross domestic product. However, official statistics likely understate productivity growth by failing to capture improvements in medical technology and treatment quality. The Health Care Satellite Account (HCSA), developed by the U.S. Bureau of Economic Analysis, addresses this gap by measuring spending by medical condition, enabling more meaningful output measurement. We present a simple framework that combines the HCSA with population health data to adjust prices and output for quality improvements. Output is defined as marginal health gains rather than service counts, consistent with prior recommendations. This approach approximates more comprehensive methods while remaining tractable. Our results suggest substantial quality-adjusted productivity growth that is largely masked in official statistics, implying a downward bias of about 1.5 percentage points per year, with a range from 0 to over 5 percentage points. Productivity gains may be larger in other high-income countries, where life expectancy has increased more and spending has grown more slowly.

* Calvin Ackley and Abe Dunn: Bureau of Economic Analysis. Eli Liebman: University of Georgia. John Romley: University of Southern California. Abe Dunn is the corresponding author (email: abe.dunn@bea.gov).
Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the Bureau of Economic Analysis. The authors would like to thank Lasanthi Fernando and Andrew Sharpe for comments.

1. Introduction

U.S. medical care spending has grown from 5 per cent of Gross Domestic Product (GDP) in the 1960s to 17 per cent in 2023.[1] Because health care is such a large share of the economy, measures of health care inflation, output, and productivity can have meaningful effects on the corresponding economy-wide measures, which are used by policymakers setting budgets, businesses making investment decisions, and other users of official statistics, including those used in monetary policy decisions. While U.S. government statistical agencies apply quality-adjustment methods to measure prices and output in other high-tech sectors of the economy, such as computers and smartphones, there is no comparable quality adjustment for the health sector. However, since the 1960s, life expectancy has increased by about nine years, much of which is thought to be driven by spending on innovations that improve health outcomes.[2] For example, more effective treatments for cardiovascular conditions, cancers, hepatitis C, HIV, and rheumatoid arthritis have extended life expectancy and improved quality of life around the world. If medical care is driving these changes in life expectancy, then many prominent papers argue that this spending is "worth it" and productivity is increasing rapidly, at least in some circumstances (Cutler and McClellan, 2001; Cutler *et al.*, 1998; Hall and Jones (2007); Murphy and Topel, 2006). However, the unadjusted measures for health care imply weak or negative productivity growth for the past several decades. According to U.S. Bureau of Labor Statistics (BLS), annual total factor productivity growth from 1990 to 2019 for health care and social assistance (NAICS 62) has fallen by 0.6 per cent per year.[3]

Currently, official statistics on health care track spending across broad service categories like hospitals, physicians, or prescription drugs, which provide limited insights into many key developments in the health sector. Namely, these aggregate figures obscure wide variation in treatments as individual diseases are associated with distinct treatment technologies. For instance, while treatments for the common cold have changed little over the past century, therapies for other conditions, such as cancers, cystic fibrosis, and multiple sclerosis, would not be recognizable from three decades ago. Grouping such diverse conditions and treatments into the same category is akin to combining the agricultural and high-tech sectors of the economy — it conceals fundamental shifts and innovations. This service-based framework has two main shortcomings. First, it fails to capture substitution patterns across medical industries in the treatment of particular conditions. For instance, shifting from

1 The growth in the share of spending by the health care sector is from BEA (https://www.bea.gov/sites/default/files/2024-04/1959-2023-BEA-Data-Applying-NHEA-Framework.xlsx).

2 See the National Center for Health Statistics (https://www.cdc.gov/nchs/data/hus/2020-2021/LExpMort.pdf).

3 In more recent years, total factor productivity growth has been flat for NAICS 62, with an annual growth of 0.16 per cent over the 2000–2019 period, but hospital and nursing care productivity (NAICS 622–623) has continued to fall. See https://www.bls.gov/productivity/tables/.

costly inpatient hospital care to outpatient visits can substantially reduce treatment costs, yet such changes are not reflected in official statistics (Aizcorbe and Nestoriak, 2011). Second, it ignores the unique technologies associated with each condition, limiting the potential to make meaningful quality adjustments.

This limitation of official statistics led to the development of the Health Care Satellite Account (HCSA) at the Bureau of Economic Analysis (BEA). Unlike traditional statistics that primarily track health care inputs (e.g., hospitals, physician offices, and prescription drugs), the HCSA redefines output to be the treatment of a condition, essentially treating each condition as its own industry (National Research Council, 2011; Cutler *et al.*, 2022). Under traditional accounts, output rises with service volume; but under the HCSA, output rises with number of patients treated. The account tracks spending, the price of treatment, and output for about 260 medical conditions (Dunn *et al.*, 2015). This framework appropriately treats hospitals, physicians, and prescription drugs as inputs into the treatment of a disease, rather than outputs, and allows for detailed insights into the diverse patterns of health care spending across different conditions. Indeed, health economists and measurement experts have long advocated for measuring the health sector output by the treatment of a condition for these reasons (Berndt *et al.*, 2000; National Research Council, 2011). However, a current limitation of the HCSA is that it does not account for quality changes brought about by improved technology.

In this article, we first derive key measurement concepts, such as quality-adjusted prices, output, and productivity. We define nominal output as expenditures on medical treatments, and apply a price index derived by Cutler *et al.* (1998) and Fisher and Shell (1972) to measure real output. This utility-based price index captures the compensating variation necessary for patients to maintain the same level of utility across periods. Consistent with Cutler *et al.* (1998) and Cutler *et al.* (2022), the utility derived from medical care includes benefits (e.g., improved health due to treatment) and costs of treatment. While the HCSA is consistent with United Nations System of National Accounts (SNA) 2025 goal of creating extended accounts that offer more detailed insights into the health care sector, the approach we take in this article deviates from the standard SNA methodology and is closer in spirit to the GDP-B framework outlined in Brynjolfsson *et al.* (2019), which aims to better capture the full benefits of economic production. More specifically, the work more closely follows the recommendations of National Research Council (2011) and Sheiner and Cutler (2024) that outline methodologies for improving measurement in the health care sector.

We then demonstrate how to construct approximate quality-adjusted estimates of prices, output, and productivity using publicly available data. Our quality-adjusted price index is constructed using estimates of the price of treatment from the HCSA, per capita health expenditures from the National Health Expenditure Accounts (NHEA), and life expectancy estimates from the National Center for Health Statistics (NCHS). In our analyses, we use aggregate (i.e., across all conditions) life expectancy improvements to quality adjust our aggregate price and output measures. The life-expectancy measure may reflect changes in population health due to non-medical factors.[4] To account for non-medical determinants of life expectancy, we incorporate information from Cutler *et al.* (2022), which develops a methodology for disentangling changes in health due to medical care from other factors.

This article builds on a much broader agenda to improve the measurement of health and health care (National Research Council, 2011; Sheiner and Malinovskaya, 2016), which includes numerous contributions in health-related literatures.[5] More specifically, this article builds on work by Cutler *et al.* (2022) and Weaver *et al.* (2022) who have developed in-depth methodologies which capture disease-specific changes in quality. However, these studies require complex methodologies and very detailed data. Our contribution is to demonstrate an approximation to these studies with readily available data sources to provide top-line analysis. In theory, this provides a starting point for more timely and transparent estimates, which is an important goal of statistical agencies. Furthermore, we extend Cutler *et al.* (2022), which focuses on the aged 65 and over population (due to data availability), to the entire population. We find that measuring the entire age distribution makes a large difference in our productivity estimates. Much of the medical care we consume at earlier ages extends or improves life beyond the age of 65. In productivity terms, many of the health production function inputs occur before 65, while many of the outputs (health improvements) accrue to those older than 65. Hence, our results suggest that a productivity measure should capture inputs made throughout the lifecycle.

Our central estimates show that quality-adjusted prices fall by about 1.3 percentage points per year relative to economy-wide inflation over the period from 2000 to 2019. This is about 1.7 percentage points below the official index for the sector (the Personal Consumption Expenditure (PCE) health price index), indicating output and productivity are understated by a similar amount.[6] Adjusting the baseline estimate of productivity from BLS to account for the

4 It is important to recognize that health outcomes are determined by numerous factors (including but not limited to behaviour and genetics), and that research suggests that medical care accounts for a fraction of the variation in health outcomes. For example, health care (or lack thereof) explains something like 10 per cent of premature mortality in the U.S.; see, among others, Schroeder (2007), Nolte and McKee (2011), and Kaplan and Milstein (2019). The social determinants of health (e.g., housing stability) are an important driver of outcomes, and have received considerable scholarly attention (Marmot and Wilkinson, 2006; Sheiham, 2009; Braveman *et al.*, 2011; Bhat *et al.*, 2023).

5 This includes literatures in cost effectiveness, health services, health policy, and health economics. For example, the seminal Dartmouth Atlas of Health Care has documented and investigated widespread variation in how health care is delivered across the United States, in turn motivating a wide range of studies in allied health fields (Wennberg *et al.*, 1996; Fisher *et al.*, 2003 a; Fisher *et al.*, 2003 b). This geographic variation has raised critical questions about why certain areas perform differently than others, leading scholars to assess, among other things, the role of low-value care (see, e.g., Chant *et al.*, 2023).

6 The BEA's PCE health price index tracks prices for health care goods and services consumed by households and reflects payments by consumers, insurers, and government programs, and is used to measure inflation in the health sector within the national accounts.

improvement in quality, we estimate productivity growth of about 1.7 per cent per year, which is substantially higher than the official estimate of 0.16 per cent per year for health care and social assistance sector.[7] The estimate is sensitive to the value placed on a healthy year of life, as well as assumptions about how much of the change in health is due to medical care versus other factors. Over a wide range of assumptions, our estimates of the quality-adjusted price index range from 0.2 per cent to -7.6 per cent per year, relative to economy-wide inflation. However, the range of spending per healthy life year saved is between $70,000 and $115,000, suggesting sizable financial cost for improved health, which should be of key interest to policymakers that must consider both limited budgets and the opportunity cost of these expenditures. Although it is important to note that the higher cost is arguably "worth it" as these amounts are substantially below the typical value placed on a healthy life year, which is typically over $150,000 (Kearsley, 2024).

The approximation presented here provides a top-line estimate which relies on more aggregate assumptions, providing an independent range of estimates to better understand the productivity of the sector. If we apply our estimates to the over-65 population, we find productivity growth estimates that are much larger than our baseline findings. The likely reason is that the share of health spending for those under 65 is relatively large compared to the improvement in health outcomes they experience. Those under 65 account for 53 per cent of the lifetime health care spending. Meanwhile, most of the health gains in life expectancy go to those 65 and over. Over the period we study, life expectancy increased 1.7 years for the 65 and over population, and 2.0 years for the full population, only an additional 0.3 years. One interpretation is that medical care spending under 65 is less productive. However, medical care spending under 65 is often an investment to lengthen one's life or improve life after 65, as in the seminal paper by Grossman (1972). For this reason, one would want to measure health care spending across the entire age distribution to measure productivity. While this is a small extension, it makes a meaningful impact in our price index and productivity estimates: our price index estimate grows -2.1 per cent per year when using only the 65 and over population versus -0.7 per cent annually when measuring the entire age distribution.[8]

As an alternative to using population health outcomes, Eggleston *et al.* (2020), Dunn *et al.* (2022), Cutler *et al.* (2022), and Dunn *et al.* (2024) use measures of clinical effectiveness of technologies from the medical literature to measure quality changes. Additional evidence based on acute health conditions, where the role of medical technology is more clear, also provides additional supporting evidence of the

7 This productivity figure is for the period 2000 to 2019, while the negative productivity growth mentioned in the first paragraph is from 1990 to 2019. More generally, the productivity for the sector is typically flat or declining slightly.

8 This number differs from the -1.3 per cent number in the previous paragraph because it assumes a VSLY of $100k to match Cutler *et al.* (2022).

productivity change (see Cutler *et al.*, 1998; Romley *et al.*, 2020; and Dauda *et al.*, 2022). We find that these alternative approaches are generally consistent with the approximation presented in this article.

The estimates presented in this article are useful to better gauge the potential bounds of mismeasurement in aggregate statistics, as well as understanding the costs of health improvement in the aggregate. However, these aggregate estimates do not reveal heterogeneous productivity differences in health care spending across populations or conditions, limiting their usefulness for health care management decisions.

While significant progress has been made in this literature, there are several important caveats. The first is that this measure of productivity is based on utility theory and consumer welfare, and is distinct from methods applied elsewhere in the accounts (see Dynan and Sheiner, 2018). The focus on welfare measurement more closely aligns with the GDP-B approach of Brynjolfsson *et al.* (2019). This approach departs somewhat from the SNA 2025 that focuses on the measurement of outputs from economic activity rather than outcomes. However, the distinction between output and outcomes blurs in the health sector as the quality-adjusted price and output depend on the expected outcome of treatments, although not (as a conceptual matter) outcomes due to non-medical factors. This distinction is important as policy implications for changes in outcome due to medical care are different for changes in outcomes due to changes in population health. A second caveat is related to the interpretation of the estimates. Evidence that productivity is improving does not necessarily imply that health spending is optimal from a welfare, budgetary, or incentive standpoint. Although the benefits appear to exceed the cost in our baseline estimates, this does not preclude the existence of alternative scenarios — such as lower spending or better health outcomes — that might generate even greater productivity gains. Additionally, policymakers may interpret these estimates differently based on their policy objectives. Some may focus primarily on short-term consumer welfare, while others might prioritize ensuring adequate incentives for firms to invest and innovate that could potentially lead to larger long-run welfare gains. Despite potentially different interpretations, providing such statistics equips policymakers with crucial information to make more informed decisions.

2. Health Care Satellite Account

Each medical condition warrants distinct treatments, underscoring the value of condition-specific data in the HCSA. The HCSA estimates begin with nationally representative survey data from the Medical Expenditure Panel Survey (MEPS), which collects detailed information on about 30,000 individuals per year, their treatment expenditures, medical conditions, and associated expenditures across all service types. While the sample size may seem large, it is actually relatively small when analyzing trends for specific conditions (see Dunn *et al.*, 2015). For this reason, the HCSA combines MEPS data with large claims databases. For the privately insured population, BEA uses the Merative™

MarketScan® Research Databases claims, which is a convenience sample of the privately insured population. For Medicare beneficiaries, the HCSA uses claims data from the Center for Medicare and Medicaid Services (CMS). For the remaining population, including Medicaid enrollees and the uninsured, the HCSA uses MEPS data. Each claims data source adds millions of enrollees and billions of claims to the estimates, and population weights are applied to maintain the representativeness of the estimates. The large sample size is necessary for capturing patterns for conditions that are costly but relatively rare in the population, such as cystic fibrosis.

In addition to tracking nominal spending by condition, the HCSA includes a price index that measures the cost of treating a condition. It is measured as the total expenditures to treat a patient over one year. For example, for heart disease, the treatment price for a patient includes all care received in a year, such as doctors' visits, labs, scans, hospital visits, and prescription drugs. Figure 1 shows trends in treatment costs for select conditions, deflated by the aggregate PCE deflator. The figure shows that the price of treating medical conditions can vary substantially over time. For example, treatment costs for rheumatoid arthritis, hepatitis, and cystic fibrosis surged as new, higher-quality drugs entered the market, followed by a sharp decline in hepatitis spending around 2015 due to increased competition among innovative drugs. In contrast, spending patterns for diabetes and heart disease were relatively flat and actually declined relative to economy-wide inflation.[9]

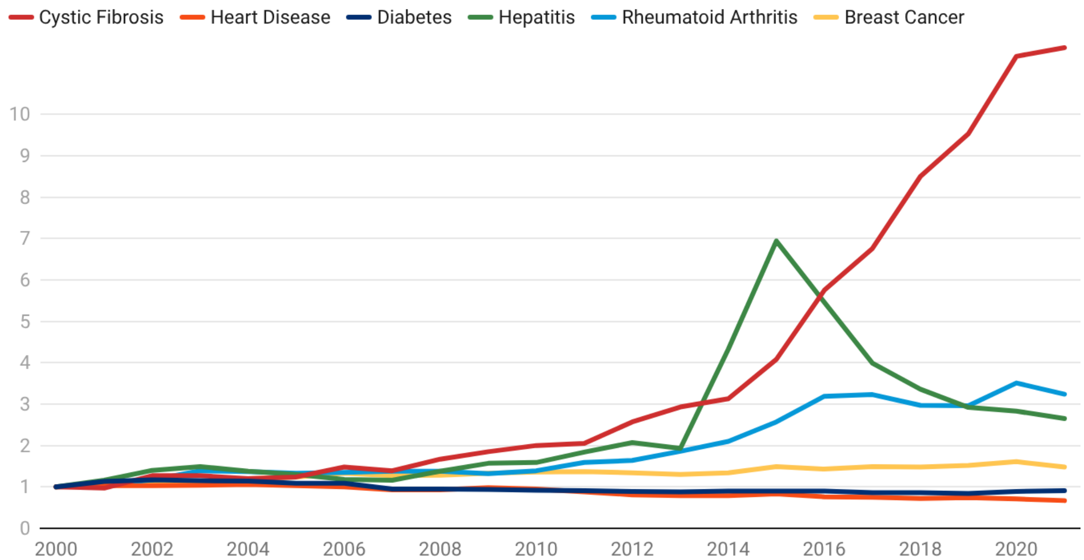
Understanding spending patterns by condition is useful for many questions that have important policy implications. What conditions account for the greatest share of spending? What conditions are driving expenditure growth, and which are slowing it down? How do these changes relate to regulations, innovations, population health, or other trends in the market? Another important use of this information is to improve measures of output and productivity to better understand the value of medical care spending. Condition-specific data make it easier to improve measures of output and productivity, as treatment technologies — and their associated costs and quality improvement — are often unique to each condition. This, in turn, allows for a more accurate assessment of the value of medical care spending. In addition, the HCSA more appropriately handles inefficient spending, relative to traditional methods. The traditional methods count output as growing as more services are provided, so inefficient spending leads to higher output. The HCSA output grows only with the number of patients, so an increase in inefficient spending per patient leads to a higher price of treatment and does not increase output.

One advantage of the condition-based price measure over traditional service-based price index measures is that it redefines the output to be the treatment of

9 This decline for heart disease coincides with a period over which many drugs to treat heart disease lost patent protection. This pattern is also consistent with Lichtenberg (2024), which finds that in the long run, pharmaceutical innovations can lead to lower costs of treatment.

Figure 1: Health Care Satellite Account Based Price Index for Select Conditions

Index value, using 2021 constant dollars



Source: Authors' calculations using disease-specific price indexes from the Bureau of Economic Analysis health care satellite account. • Created with Datawrapper

a condition, which better handles substitution patterns across different types of inputs, as highlighted by Aizcorbe and Nestoriak (2012).[10] Aizcorbe and Nestoriak (2012) demonstrate that shifts could lead to cost savings, leading condition-based price indexes to grow more slowly than traditional indexes (e.g., shift from expensive inpatient services to outpatient services).[11]

Alternatively, if more expensive new technologies are used in treatment, this could lead the condition-based index to rise more quickly than the traditional PCE health price measure. Specifically, if an expensive new treatment enters the market in year 2, it will not be added to the PCE index in tracking price changes from year 1 to year 2, as it tracks a fixed basket of goods and services and excludes the new treatment. In contrast, the HCSA will increase in year 2 when the new treatment replaces older, cheaper treatments. If the new and higher price technology is also of higher quality, both indexes will be biased, but the HCSA may appear to have a larger bias, as the new technology leads to a bigger increase in the index, relative to the PCE health measure.

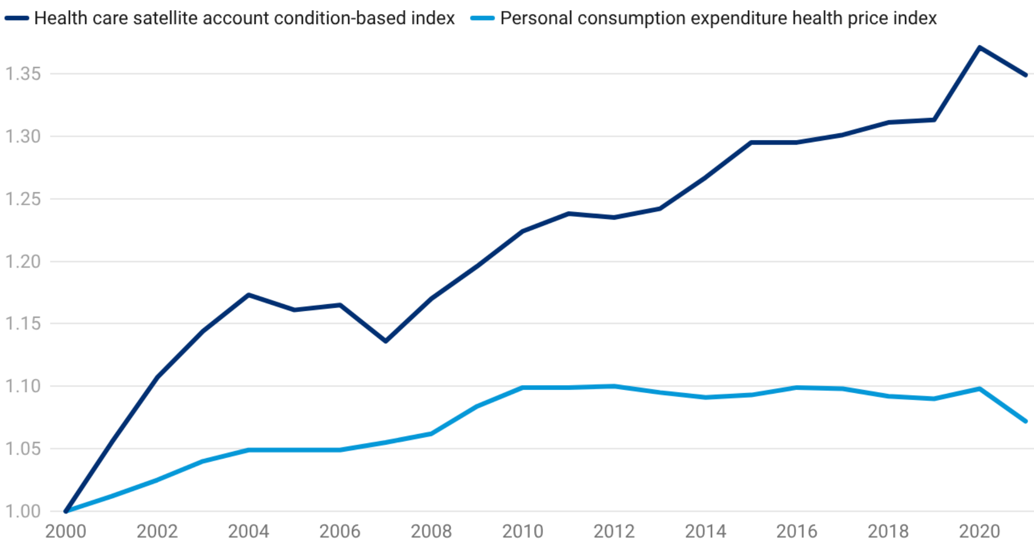
Theoretically, either effect (e.g., shifting services to cheaper settings, or new innovative treatments raising costs) could dominate, but as shown in Figure 2, the

10 Aizcorbe and Nestoriak (2012) is the first paper to construct condition-based estimates using claims data, providing an important foundation for the development of the HCSA.

11 In this case, if there is no quality change, then the condition-based index would more accurately capture inflation, relative to the official PCE health price index, which measures the cost of a basket of specific goods and services year over year.

Figure 2: Condition-Based Aggregate Price Index and Personal Consumption Expenditure for Health Price Index

Indexed to 1 in 2000



Source: Authors' calculations using data from Bureau of Economic Analysis health care satellite account (as of December 2-23, 2025), personal consumption expenditure health price index (as of June 27, 2025), based on national income and product accounts estimates. • Created with Datawrapper

condition-based price index tends to grow faster than the traditional PCE health price index, highlighting the role of technology likely driving up the cost of treatment. In fact, the condition-based price index grows faster in many cases because of the adoption of newer and costlier technologies, such as for the treatment of several of the conditions shown in Figure 1. At the same time, micro evidence shows shifts toward higher quality treatments. Dunn *et al.* (2024) examine innovations for 13 health conditions and show that there is a tendency for consumers to gravitate toward higher quality new treatments, even if those treatments are substantially more expensive. Chandra *et al.* (2016) show that patients gravitate toward higher quality hospitals over time, even for acute health conditions. This evidence highlights the need to properly quality adjust condition-based price indexes.

3. Cross-Country Differences

This article focuses on the U.S. experience, but there is broad international interest in this topic due to both the importance of health and the growing share of government budgets across economies. Cross-country comparisons further highlight the link between economic measures and health outcomes. Deaton and Schreyer (2022) show a strong positive correlation between national accounts aggregates — particularly actual individual consumption — and life expectancy. While this relationship is evident, what remains largely unmeasured is how such health outcomes connect to the output and productivity of the health care sector, despite their clear significance for both economic performance and overall well-being. A key challenge is isolating the contribution of medical care from broader non-medical determinants of health

(Sharpe *et al.*, 2007; Schreyer, 2012; National Research Council, 2011; and Sheiner and Cutler, 2024).

One fact that stands out from these comparisons is that the U.S. spends substantially more on health care than most other economies, yet health outcomes measured by life expectancy and some other metrics are substantially worse (Peter G. Peterson Foundation, 2020). There are a variety of theories regarding what drives these differences, including administrative costs, pricing, adoption of high cost technologies, heterogeneity in treatment across the population, and differences in underlying population health (Garber and Skinner, 2008; Cutler and Ly, 2011; Chandra and Skinner, 2012; and Einav and Finkelstein, 2023). While aggregate statistics suggest that the U.S. health sector is less efficient, several studies suggest that higher obesity rates may contribute substantially to lower life expectancy, complicating these comparisons (Preston and Stokes, 2011; and Mokdad *et al.*, 2024). To compare the productivity of health systems over time or across countries, it is necessary to account for the differences in population health affecting both outcomes and spending to better isolate the effects of the health care sector on health outcomes and spending. In this article, we focus on the productivity in the United States over time.

4. Literature

This section highlights key insights relevant to measuring productivity in the health sector. For a more comprehensive review of the literature, see Hall (2017), Sheiner and Malinovskaya (2016), and Sheiner and Cutler (2024).

This article focuses on a utility-based measure of productivity, where we show the change in output per treatment depends on improved health. Alternative methods tend to greatly understate the value of health improvements, as highlighted in several recent papers (Sheiner and Malinovskaya, 2016; Dauda *et al.*, 2022; and Dunn *et al.*, 2022). Additionally, hedonics used in other industries, such as computers or smartphones, are poorly suited to the health care sector, where there are numerous market distortions and consumers rarely face the marginal cost of treatment (Berndt *et al.*, 2000).

There is broad evidence that the quality of treatments is improving over time, leading to improved health outcomes. This evidence comes from studies using population health to measure quality, analysis based on the clinical literature, and studies examining acute health conditions (see Sheiner and Cutler, 2024, for a recent overview). Outside of the United States, recent evidence from Park *et al.* (2025) shows large quality improvements in South Korea, likely attributable to improvements in the health sector. Eggleston *et al.* (2020) show quality-adjusted prices falling for diabetes care across several countries. While many papers in this literature focus on specific health conditions, the goal of this article is to estimate an aggregate quality-adjusted price index for the United States.

The methodology used to derive a quality-adjusted price and productivity relies on the dollar value placed on improved health (see Viscusi, 2020, for a review). Recent estimates of a value of a life in the U.S. range from \$6 million to \$20 million

(Kearsley, 2024), but the exact magnitude is important for assessing the productivity gains in the sector. Using the dollar value of health from Kearsley (2024), the value for a year of health is $150,000 or more.[12] Our analysis follows Cutler *et al.* (2022) and uses a value of $100,000 for comparability. However, given the uncertainty and higher range implied by the literature, we also provide estimates using $150,000 and $250,000 per statistical life year.

5. Measuring Output and Productivity in the Health Care Sector

Understanding both the benefits and the costs of treatment is fundamental for measuring the productivity of the health sector, but numerous distortions in health care markets complicate standard approaches for measuring quality change. Seminal work by Cutler *et al.* (1998) addresses this issue by developing a framework based on utility theory that is robust to market distortions in the health care sector. This section presents basic formulas for price, output, and productivity derived from this utility-based framework.

5.1 Quality-Adjusted Price Index

Following Cutler *et al.* (1998), Sheiner and Malinovskaya (2016), and Dauda *et al.* (2022), let $P^j_{0,t}$ be the quality-adjusted price index for a condition j from time period 0 to t. We apply a Laspeyres-type price index that uses a compensating variation welfare formula to adjust for quality.[13] The quality-adjusted price index measures the growth in treatment expenditure from period 0 that would be necessary to maintain the same level of utility in period t, holding technology constant. Let S^j_t be the lifetime expenditure for treatment of condition j at time t.[14] Let H^j_t be the incremental amount of health produced from treatment at time t for disease j. H^j_t is often measured in terms of quality-adjusted life years (QALYs) that account for both the number of years alive and the quality of those years, where one QALY is one year of life in perfect health.[15] Importantly, H^j_t is not the actual health outcome, but the amount of health due to treatment and not other factors (e.g., diet or exercise). Both S^j_t and H^j_t are typically risk-adjusted to account for the age and health conditions of the patient, so that changes in expenditures and health outcomes are the changes from medical care, and not changes from other factors. In

12 Recent estimates in the literature, such as Cutler *et al.* (2022), do not apply discounting to spending or health improvements, which would lead to a statistical life year of $150,000 for a low-end estimate based on figures from Kearsley (2024).

13 A similar Paasche-type price index could also be applied, as shown in Dauda *et al.* (2022).

14 Spending needs to be measured in expected lifetime units to match the QALY measurement, which is in terms of the gain in health over a lifetime.

15 For instance, a health condition that leads to a disability will reduce the QALY of a patient, even if the life expectancy does not change.

order to quality adjust the price of treatment, a value is needed to convert units of health into dollars. Typically, researchers apply measures from external studies that derive estimates of the value of a statistical life year, as discussed previously. Here we let the dollar value of a statistical life year be represented as $\$VQALY$ (Value of a Quality-Adjusted Life Year), and we consider the range from $100,000 to $250,000. The change in quality of treatment is then $\Delta H_t^j = H_t^j - H_0^j$, so the dollar value of the change is: $\$VQALY \cdot \Delta H_t^j$.

We follow Dauda *et al.* (2022) and Sheiner and Malinovskaya (2016) to compute a quality-adjusted price $P_{0,t}^j$ for condition j at time t. The price index is an index of the growth in price from the base period 0 to the end of our sample t. Intuitively, the quality-adjusted price asks how much spending in period 0 would be needed to deliver the same utility as treatment in period t. In order to hold technology constant, the quality improvements must be subtracted from the unadjusted treatment price, S_t^j, in period t.

$$P_{0,t}^j = \frac{S_t^j - \$VQALY\Delta H_t^j}{S_0^j} \tag{1}$$

$$= \frac{S_t^j}{S_0^j} - \frac{\$VQALY\Delta H_t^j}{S_0^j} \tag{2}$$

The second row of the index shows that without any change in the quality of treatment, the price index measure would be a measure of the unadjusted price of treatment, $\frac{S_t^j}{S_0^j}$, which is the price index in the HCSA. Quality improvements lead to reductions in the quality-adjusted price of treatment through the adjustment term in the numerator, $\$VQALY\Delta H_t^j$. The derivation is in the appendix, but this formulation is intuitive. Consider the example of the price change from the introduction of the drug Sovaldi in 2014 used to treat Hepatitis C discussed in Dunn *et al.* (2022). They consider the price change compared to the prior drug interferon. Based on cost-effectiveness studies, the price of Sovaldi was $105,488, while the price of interferon was $81,211. The Quality-Adjusted-Life-Years (QALYs) from treatment is 9.4 for Sovaldi and 8.28 for interferon. As an illustrative example, they place a relatively low value on the QALY of just $50,000.

16 The value is derived as: 0.61 = ($105,000 - $50,000*(9.4-8.28))/$81,000.

17 Note that sufficiently large quality improvements can potentially yield a negative price index. There are three primary ways to address this issue. First, Dauda *et al.* (2022) apply chaining, which effectively scales down incremental quality gains at each step relative to costs. A second approach is to expand on the conditions over which quality adjustments are applied. By increasing total treatment expenditures, this helps ensure that the resulting price index remains positive. If chaining is not feasible and the application requires condition-specific price indexes, then a common solution is to use a reservation price index. Dunn *et al.* (2022) and Ackley *et al.* (2026), building on Trajtenberg (1990), use a reservation price index of the form: $P_{0,t}^j = \frac{S_t^j}{S_0^j + \frac{\$VQALY\Delta H_t^j}{S_0^j}}$, where the quality adjustment enters the denominator. This index ensures that improvements in technology do not generate negative prices because the denominator represents the reservation price that leaves individuals in the base period indifferent between the new and prior technologies. While all three approaches are theoret-

Based on these estimates and the utility-based price index formula, the price index falls by 39 per cent.[16] Intuitively, the average person is receiving \$56,000 = \$50,000*(9.4-8.28) of value for the health they are purchasing, at an incremental cost of \$24,000 = \$105,000 - \$81,000, hence quality adjusted prices are falling.[17]

One seemingly intuitive alternative price adjustment is to scale the price by the quality change, for example, $P_t = \frac{S_t}{S_0}\frac{H_0}{H_t}$. Intuitively, one can view this as the change in the price of a QALY. However, as pointed out in a number of recent papers (Sheiner, 2016; Dunn *et al.*, 2022; and Dauda *et al.*, 2022), scaling price changes by the growth in quality of treatment does not account for the value of a QALY. Here is a stylized example to highlight this. Suppose an individual with high cholesterol in 1980 has a baseline life expectancy of 10 years and their cholesterol medication cost \$1,000. By 2015, statins have entered the market and gone off patent. Suppose that the same individual would now have 15 years of life expectancy, and their statins cost \$2,000. This index suggests that quality adjusted prices have risen by 33 per cent ($\frac{\$2{,}000}{\$1{,}000} \times \frac{10}{15}$), placing an implicit value of the health gain of around \$670.[18] However, most economists would argue that a 5-year increase in life expectancy is worth more than several hundred dollars. Our measure differs because it places an explicit value on a year of life, while the price-per-QALY approach is orthogonal to changes in the value of a life year, as that term cancels out of the ratio H_0/H_1.

Others in the literature have pointed out the differences between these two intuitive ways of quality adjusting health spending (Sheiner, 2016; Dunn *et al.*, 2022; and Dauda *et al.*, 2022). Specifically, Sheiner and Malinovskaya (2016) points out that if one assumes that the price-per-QALY is roughly the VSLY, then our indexes are the same.[19] While this type of assumption is valid in many contexts, it is often violated in health care markets. Some very cheap technologies, like generic drugs, can provide a lot of health benefits at very low costs. But, in many cases, it is not possible to buy more health, even with very high levels of expenditure, due to technological constraints.

5.2 Growth in Real Output

Real output per case (i.e., per patient), Y_t^j, is obtained by dividing spending per case by the quality-adjusted price index.[20] Using equation 1, we have $Y_t^j = \frac{S_t^j}{P_{0,t}^j} = \frac{S_t^j \cdot S_0^j}{S_t^j - \$VQALY\Delta H_t^j}$. Let the number of cases be N_t^j in period t and N_0^j in period 0.

ically valid, they differ in the underlying baseline utility and technology being held fixed. The approximation used in this article follows the third approach.

18 The unadjusted price index is 2 (i.e., $\$2,000/\$1,000$), while the adjusted price index that scales the price by the quality change is 1.33 (i.e., $\$2,000/\$1,000 \cdot 10/15 = 1.333$). If this adjustment is capturing the correct reservation price, then the expenditures necessary to make an individual indifferent to the new technology is just \$1,333, so the dollar adjustment is \$670 for five additional years of life.

19 The price per QALY formula is: $P_t = \frac{S_t}{S_0}\frac{H_0}{H_t} \cong \frac{S_t}{S_0}\left(1 - \frac{\Delta H}{H_0}\right) = \frac{S_t}{S_0} - \frac{\Delta H}{S_0}\frac{S_t}{H_0}$. If $\$VSLY = \frac{S_t}{H_0}$, then this formula matches equation 1.

20 As applied in the U.S. National Accounts, deflators are applied to obtain real output.

Total real output in period t is $N_t^j \cdot Y_t^j$ and the real output in the base period is $N_0^j \cdot Y_0^j = N_0^j \cdot S_0^j$. Therefore, the index of quality-adjusted total real output growth, reflecting real personal health care consumption, is:

$$\frac{N_t^j \cdot Y_t^j}{N_0^j \cdot Y_0^j} = \frac{N_t^j \cdot S_t^j}{N_0^j \cdot S_0^j \cdot P_{0,t}^j} = \frac{N_t^j \cdot S_t^j}{N_0^j \cdot (S_t^j - \$VQALY\Delta H_t^j)} \tag{3}$$

The index measure of real output growth is an increasing function of the health gained from the treatment: ΔH_t^j that is scaled to a dollar value, times the growth in the number of cases. If the quality does not change, then the output growth does not depend on the amount spent on treatment, but only the number of treatments. If quality does change, then an adjustment is needed to convert output per case in the base period to the output per case in period t. It is worth emphasizing that this is a stark difference from more traditional output measures. Holding the number of cases $N_t = N_0$ constant, then the quality-adjusted output measure only increases when health from treatment improves, while traditional measures increase as more goods and services are provided (e.g., doctors visits or prescription drugs).

The growth in output is *not* proportional to the growth in health. It depends on the increase in health relative to the amount spent on the treatment.[21] For countries that do not measure output growth by deflating expenditures, the adjustment could be applied directly to an output per case growth measure in equation 3.[22]

5.3 Productivity

The productivity change is determined by the growth in real output relative to the growth in inputs. Let $C_t^{N,j}$ be the nominal (input) cost per case, then to obtain the real cost per case, $C_t^{R,j}$, we divide by an input price index. The cost includes the associated cost of inputs (e.g., capital, labour, and materials), and may differ from S_t^j if more output can be produced with the same level of costs. The index of real input growth is then $\frac{C_t^{R,j}}{C_0^{R,j}}$. Productivity growth is the growth in real output divided by the growth in real input:

$$ProductivityIndex_{0,t}^j = \frac{N_t Y_t^j}{N_0 Y_0^j} \Big/ \frac{N_t C_t^{R,j}}{N_0 C_0^{R,j}} = \frac{Y_t^j}{Y_0^j} \Big/ \frac{C_t^{R,j}}{C_0^{R,j}}$$

Inserting the formula for growth in real output and growth in costs we have:

21 For example, suppose a treatment cost \$20,000 in period t, but the health produced from a treatment changed from 0.5 QALY to 0.6 QALY and $\$VQALY = \$100,000$, then based on the formula the output doubles, even though QALY increased by 20 per cent. The intuition is straightforward, as relative to the \$20,000 in output received in period t (the numerator), the output is the base period is \$10,000 in value after accounting for the lower quality of treatment received in the base period. Patients are getting twice the output per dollar spent.

22 Note that volume measures applied internationally are often counts for particular places of service (e.g., hospital volume or physician visit volume), whereas the volume measure here is the number of patients treated for the condition, across all places of service.

$$ProductivityIndex^j_{0,t} = \frac{\frac{S^j_t}{S^j_t - \$VQALY\Delta H^j_t}}{\frac{C^{R,j}_t}{C^{R,j}_0}} \tag{4}$$

Equation 4 shows that the quality adjustment enters through the output price index, so one way to adjust the official measure of productivity is to apply an adjustment to the output price index (see Dunn *et al.*, 2022), which is the approach we take in the empirical section below. More precisely, we are essentially changing the deflator applied to output by multiplying the official growth in real output by the ratio of the corresponding official deflator, divided by the quality-adjusted price index.

6. Empirical Evidence

6.1 Estimates Based on the Health Care Satellite Account

This section reports estimates of the quality-adjusted price index and productivity index, based on the formulas developed in the previous section. To avoid distortions from COVID-19, we focus our estimates on the period 2000–2019.

We present an aggregate quality-adjusted price index that is a simplified version of Cutler *et al.* (2022). At an aggregate level, the analysis can be simplified as it avoids explicit allocation across conditions, where it may be challenging to match spending to health outcomes. For instance, suppose a patient dies with renal failure and heart disease, two serious conditions that often appear together. In this case, it is challenging to attribute this change in health across these conditions, as both likely contributed to the death. A similar allocation is necessary for spending across the two conditions. To address this issue, Cutler *et al.* (2022) use a complex propensity score methodology that requires detailed micro data on conditions and health outcomes. However, we can observe changes in aggregate health, which avoids the complexities of attribution, and is sufficient for aggregate measures of productivity.

Key inputs are shown in the top panel of Table 1. All inputs are deflated by the aggregate PCE index and all of the expenditure computations in Table 1 are lifetime computations from birth. The first row shows that the PCE health index grows faster than overall inflation by 0.45 per cent per year. The condition-based price index increases considerably faster, as discussed previously, 1.4 per cent per year faster than economy-wide inflation, consistent with Figure 2. The next three rows show the life expectancy for 2000 and 2019, along with the gain in life expectancy. Life expectancy increased by roughly 2 years during our sample period. The base period lifetime spending in 2000 is \$512,877.[23] Multiplying the HCSA condition-based price index by the base

23 This is computed using estimates of spending per capita by age from CMS combined with life expectancy tables for the year 2000 from the Center for Disease Control National Vital Statistics Report. In principle one could construct this again in 2019, but an advantage of the HCSA is that it holds the prevalence of conditions fixed across time, essentially holding the health of the population constant.

Table 1: Quality-Adjusted Price Changes Based on the Health Care Satellite Account

Inputs into quality-adjusted price index

Description	Value
Personal consumption expenditure health price index annual growth	0.45%
Condition-based price index annual growth	1.40%
Life expectancy 2000	76.8
Life expectancy 2019	78.8
Life expectancy gain in year (2000-2019)	2.0
Estimated lifetime spending in 2000	$512,877
Hypothetical risk-adjusted lifetime spending in 2019	$673,483
Change in lifetime spending 2000 to 2019	$160,607

Annual change in quality-adjusted price index

Scenario	Change in quality-adjusted life years	Assuming VSLY = $100,000	Assuming VSLY = $150,000	Assuming VSLY = $250,000	Implied health spending cost per change in QALY
(1) Life expectancy gains = QALY (baseline)	2.0	-0.42%	-1.66%	-5.55%	$80,303
(2) QALY reduced 30 per cent from baseline	1.4	0.21%	-0.53%	-2.40%	$114,719
(3) Worsening population health gains, scenario 2 understated by 30 per cent	1.8	-0.22%	-1.29%	-4.39%	$88,245
(4) Worsening population health gains, scenario 2 understated by 60 per cent	2.2	-0.69%	-2.18%	-7.63%	$71,669

Note: QALY = Quality-adjusted life year; VSLY = Value of statistical life year.

Source: Authors' calculations using the Bureau of Economic Analysis health care satellite account. • Created with Datawrapper period lifetime spending implies hypothetical lifetime spending of $673,483 in 2019, or an increase of $160,607 in lifetime spending for a population of similar health in 2019, relative to 2000.

Using this information, we form a variety of estimates to better gauge the value of medical care spending. We start with a simple but important benchmark, Scenario (1), where we assume all of the life expectancy change may be attributable to the health care sector. Scenario (1) on the bottom panel of Table 1 also ignores disability and assumes that the QALY gains equal the life expectancy change. In this scenario, spending change per year of life expectancy gained equals $80,303, because life expectancy rose by 2 years and spending by $160,607. For a $VQALY of $100,000 this implies an annual quality-adjusted price decline of 0.42 per cent, relative to economy-wide inflation.[24] Using $VQALY from $150,000 or $250,000, which is a range more consistent with recent estimates from Kearsley (2024), we find the price index falling substantially faster, from about 1.66 per cent to 5.55 per cent per year.

In the next row (Scenario 2), we reduce the growth in quality-adjusted life expectancy, reflecting the fact that the additional life years that individuals gain may be in less than perfect health. We follow Cutler *et al.* (2022) who find that QALY gains are 30 per cent less than the life expectancy gains over the period of study. This adjustment reduces the value of medical care spending substantially. With a VSLY of $100,000, we find that quality-adjusted prices rise by 0.21 per cent annually, rather than fall by 0.42 per cent in Scenario (1).

Finally, in the last two rows, we assume that the underlying health of the population has worsened for non-medical reasons, implying the observed change in health due to medical care is understated. This assumption is consistent with trends in underlying population health in the U.S. Due to rising obesity rates and drug abuse (see Mokdad *et al.*, 2024), it may be argued that the underlying population health has declined over time due to non-medical factors. For the 65 and over population, Cutler *et al.* (2022) decompose how much of the change in health is due to medical and non-medical factors. Over the period from 1999 to 2012, they find that the gains in health due to medical care are larger than gains in health generally, as obesity, among other factors, has reduced health. Because of this, they find changes in quality-adjusted life expectancy understate the impact of medical care by about 60 per cent for the

24

$$
\begin{aligned}
P &= \frac{S_t}{S_0} - \frac{\$VQALY\Delta H_t}{S_0} \\
&= \frac{\$673,483}{\$512,877} - \frac{\$100,000 \cdot (2 \text{ Years})}{\$512,877} \\
&= 0.92
\end{aligned}
$$

The value 0.92 is the index value over the entire time period. Annualized over the 19 years, the value is 0.9958, so the annualized price change is -0.42%.

65 and over population. Given that we are analyzing spending and outcomes at birth, this estimate from Cutler *et al.* (1998) may not apply directly. Therefore, we analyze a range of estimates. In the last two scenarios (3 and 4) we assume gains in health from medical care are understated by 30 per cent or 60 per cent, relative to Scenario 2. We choose 30 per cent as our preferred estimate, in addition to 60 per cent, as we are analyzing a younger, potentially healthier population. Based on these adjustments, the central estimates show the quality-adjusted price index falling for most scenarios. Interestingly, across all four scenarios the quality-adjusted price index falls below the PCE health price index growth of 0.45 per cent per year over this period, implying a potential bias in the official price measure in the sector across the range of scenarios. The bias is the difference between the current PCE health price index and the quality-adjusted price index. For example, for our preferred Scenario (3) assuming the VSLY of $150,000, correcting for the bias would decrease the price index by 1.74 per cent (= 0.45% - -1.29%) per year. As the price indexes are sensitive to the VSLY, the last column reports the change in spending per life year saved as a measure of the growth in expenditures per quality improvement. In all scenarios, the spending per QALY increase is over $70,000, which may be of interest to policymakers considering the opportunity costs of these expenditures.

To check whether our approximation methodology provides plausible estimates, we compare our results to the much more in-depth estimates from Cutler *et al.* (2022). One important difference between our papers is that Cutler *et al.* (2022) focuses solely on the 65 and over population, while all of our calculations in Table 1 are lifetime calculations from birth. To make our results comparable, we recalculate life expectancy at 65 and lifetime spending at 65 in 2000, using age specific death rates and health care spending.[25] The HCSA does not currently separate spending for those above or below 65, so we assume spending growth is similar for these two populations, which is an arguably strong assumption that we discuss below. After making these adjustments, the base period lifetime spending of a 65 year old is $295,738, the change in lifetime spending is $92,610, and the change in life expectancy is 1.7 years. This corresponds to a quality-adjusted price index decline of 2.09 per cent in Scenario 4, with a VSLY of $100,000. With the same assumptions regarding the VSLY and the share of health improvements due to medical care, we apply our formula to estimates from Cutler *et al.* (2022) and find that the implied quality-adjusted price index falls by about 3.5 per cent per year.[26] This is a bit lower than our estimates, but the over-65 population in the U.S. is insured by Medicare, which

25 See the appendix for specific calculations for the over-65 age group.

26 Specifically, we use their estimate of the change in lifetime spending and QALYs from Table 5 of Cutler *et al.* (2022). We compute base period lifetime spending using our methodology (but for 1999, which is the beginning of their sample period), as they do not report this number. We adjust all their numbers to be in 2019 dollars, rather than 2010 dollars.

has regulated prices that grow less quickly over our study period. Private insurance is the most common insurance for those under age 65.[27] If we factor in the lower price growth in the Medicare population, which grows over 1 percentage point slower than private insurers, then our estimate matches closely to those of Cutler *et al.* (2022).[28]

There are large differences in the quality-adjusted price index for the full population analyzed in Table 1, relative to the 65 and older population analyzed in Cutler *et al.* (2022). The reason for this large difference is that the health gains go primarily to those over 65, but the lifetime health spending is shared roughly evenly between the under and over-65 populations, so it appears that the value of medical care spending for the over-65 population is quite high. More precisely, about 85 per cent of the improvement in life expectancy is occurring for those older than 65, while about 53 per cent of the spending growth occurs below 65. One possibility is that the increase in spending is much more effective for the over-65 population. However, we think that the more likely reason for this difference is that health is an investment good, as in Grossman (1972), so that the spending below age-65 leads to better health outcomes post-65, highlighting the potential importance of studying the dynamics of the full population.

These estimates have implications for productivity measurement. To understand the implications, we adjust existing productivity estimates for health care from BLS. Analogous to equation 4, we can write the BLS Multifactor Productivity index as:

$$ProductivityIndex^{BLS}_{0,t} = \frac{\frac{\frac{N_t \cdot S_t}{N_0 \cdot S_0}}{\text{Output Price Deflator}_{0,t}}}{\frac{N_t C^R_t}{N_0 C^R_0}} \tag{5}$$

where $N_t \cdot S_t$ is nominal output for health care in period t, $N_0 \cdot S_0$ is nominal output in period 0, and Output Price $\text{Deflator}_{0,t}$ is the BLS price deflator for health care. To account for the changes discussed above (e.g., quality adjustment and defining the output as the treatment of a condition), one could replace the BLS price deflator with our quality-adjusted price index by multiplying the productivity measure in equation 5 by $\frac{\text{Output Price Deflator}_{0,t}}{P_{0,t}}$.

The top panel of Table 2 shows the inputs to this calculation. The BLS multifactor productivity change for the sector is relatively flat, with an increase of just 0.16 per cent per year. The output price deflator grows by 0.24 per cent per year relative to economy-wide inflation. Note that the BLS price measure we use is Health Care and Social Assistance (NAICS 62), which varies from the BEA measure of PCE Health. This sector includes hospitals and

27 Based on estimates from the Bureau of Labor Statistics for hospitals, prices for individuals with private insurance grew by about 4 per cent per year, while prices for individuals with Medicare insurance grew 2.5 per cent per year.

28 Specifically, as an alternative we lower the disease specific price by half the price difference between private and Medicare price growth (about 0.7 percentage points a year) to account for the slower growth of Medicare prices, and we find the associated quality-adjusted price falls by 3.5 per cent. After these adjustments, the similarity between our estimates provides some evidence that our methodology to approximate a quality-adjusted price index is reasonable. Details of this calculation are shown in the appendix, Section 8.3.

Table 2: Quality-Adjusted Productivity Changes Based on the Health Care Satellite Account

Inputs into quality-adjusted productivity index

Description	Value
BLS multifactor productivity health and social assistance	0.16%
BLS output price deflator health and social assistance	0.24%

Annual change in quality-adjusted productivity index

Scenario	Assuming VSLY = $100,000	Assuming VSLY = $150,000	Assuming VSLY = $250,000
(1) Life expectancy gains = QALY (baseline)	0.82%	2.06%	5.95%
(2) QALY reduced 30 per cent from baseline	0.19%	0.93%	2.80%
(3) Worsening population health gains, scenario 2 understated by 30 per cent	0.62%	1.69%	4.79%
(4) Worsening population health gains, scenario 2 understated by 60 per cent	1.09%	2.58%	8.03%

Note: QALY = Quality-adjusted life year; VSLY = Value of statistical life year.

Source: Authors' calculations using the Bureau of Economic Analysis health care satellite account. • Created with Datawrapper

ambulatory care, but unlike PCE Health, it excludes pharmaceuticals and includes the category of social assistance.[29] These differences are among the reasons that the BLS price deflator grows slower, 0.24 per cent versus 0.45 per cent, relative to the PCE health price index.

Table 2 shows how this adjustment impacts productivity measures. In Scenario (3), assuming a VSLY of $150,000, we calculate that the annual productivity growth is 1.69 per cent from 2000 to 2019, about a ten-fold difference in productivity. This implies a bias of 1.53 per cent (= 1.69% - 0.16%) per year. One can see that with this adjustment, the productivity estimate mirrors the quality-adjusted price index change. For example, in Scenario (4), assuming a VSLY of $100,000, our productivity estimate is 1.09 per cent per year, which

29 However, social assistance accounts for less than 10 per cent of this category, so it primarily reflects productivity of the health sector.

implies a bias of 0.93 per cent (= 1.09% - 0.16%) per year, or a seven-fold difference in productivity.

Limitations: While we view this approach as providing a useful estimate for a quality-adjusted price, there are some important limitations. Most importantly, the paper relies on external information on population health and assumptions regarding how much of the population health change is due to medical care or other factors. The recent work by Cutler *et al.* (2022) provides some guidance suggesting that population health in the U.S. has likely worsened in recent years. The increase in disease prevalence in the HCSA is also indicative of worsening population health. Consistent with these estimates, Romley *et al.* (2020) examine eight health conditions around this time period and finds a health status index that worsens for seven out of eight conditions studied. There are also challenges with measuring health improvements based on population life-expectancy. For example, life-expectancy measures may not capture more complex dynamics, such as treatments that only have effects on health several years into the future.

Another important limitation is that our approach does not address the issue of clinical risk factors, like high cholesterol or diabetes, that can lead to more serious conditions, such as heart disease, that can lead to morbidity and death. If hypertension and high cholesterol increase in the population, this can lead to more heart disease cases. An important contribution of Cutler *et al.* (2022) is to develop an accounting system to reallocate spending and health changes from direct health conditions to risk factors like hypertension and high cholesterol. This more appropriately lines up spending and associated outcomes for each condition. Our analysis attempts to address this limitation by using health information from Cutler *et al.* (2022).

Another limitation of the analysis is its focus on aggregate estimates. We do not have health information at the condition level, which can be informative regarding the quality-adjusted prices and productivity of different treatments. More disaggregated information, by condition, geography, and demography, is necessary to better understand why productivity is changing. However, the benefit of our approach is that it can more quickly be applied to provide a more timely top-line measure. The estimate presented here is just an approximation, and more information is needed to improve the accuracy of these estimates. While comprehensive estimates are challenging due to the vast data required, Cutler *et al.* (2022) has demonstrated the feasibility for the over-65 population and the Institute of Health Metrics and Evaluation also demonstrates the feasibility of producing comprehensive estimates by providing disease outcome information for 204 countries (Vos *et al.*, 2020).

6.2 Additional Evidence

A key limitation of the HCSA-based approach presented here and by Cutler *et al.* (2022) is that there may be numerous unobserved risk factors affecting the health of individuals with a particular condition. For instance, the health of a person with diabetes may be very different across individuals and over time, but individuals with a more severe case are often coded the same as someone with a less severe case. Detailed risk factors of the person may be difficult to account for using only population health data. There are a few complementary approaches that have been taken that generally support the findings presented here.

One approach is to produce estimates based on the expected quality from disease models, which use evidence in the clinical literature to predict health outputs based on information about health inputs and patient health (National Research Council, 2011). Cutler *et al.* (2022) introduces an alternative analysis using a disease model for cardiovascular treatment. This approach does not rely on population health outcomes and finds results consistent with the population-based approach for cardiovascular conditions. Eggleston *et al.* (2020) use disease models to examine the value of treatment for diabetes patients in four countries, and find the benefit of health improvement typically exceeds the cost.

Related to the disease model approach, Dunn *et al.* (2024) study 13 health conditions and use information on quality from the medical literature from the Tufts' Cost Effectiveness Analysis Registry (CEAR). Specifically, they combine quality information from clinical studies with actual spending and treatment information using micro claims data. Many of the 13 conditions studied have expensive treatments that are patent protected, and quality-adjusted prices rise over the study period, highlighting that quality-adjusted prices do not necessarily fall. However, they project that in the long run, when drugs lose patent protection, quality-adjusted prices fall substantially for 12 of the 13 conditions.[30] A potential advantage of the approach taken in these papers is that quality is not measured based on observed outcomes; it is based on the clinical literature that provides a measure of predicted health outcome. A disadvantage of this approach is that it does not directly explain health outcomes observed in the population. Moreover, as Hall (2017) points out, it may be challenging to model every separate condition and associated outcome. See Hall (2017) and Sheiner and Cutler (2024) for more detailed discussions of other papers in this literature.

Focusing on acute health conditions is another approach to help address this issue, as the acute health event itself and associated diagnosis codes provide key information about patient health. Due to the severity of these conditions, the outcome and the spending soon after the event can more readily be attributed to the health sector. This literature typically finds quality-adjusted prices declining rapidly (Cutler *et al.*, 1998, for heart attacks; Dauda *et al.*, 2022, for heart attacks, pneumonia, and heart failure). Examining eight health conditions, Romley *et al.* (2020) finds evidence of broad productivity improvement.

30 Similar to this study, Dunn *et al.* (2022) use available information from the medical literature to approximate a quality-adjusted price, but at a more aggregate level. Specifically, they use information on advances in medical technologies from cost-effectiveness literature, proxies for the diffusion of medical technology, and combine this information on the cost of treatment by medical condition from the HCSA. They also find quality-adjusted prices declining, similar to the population-based approach discussed in this article.

7. Conclusion

Accurately measuring productivity in the health care sector is essential for understanding the value of medical care spending, informing policy, and interpreting broader economic trends. However, traditional approaches do not adequately account for improvements in health outcomes and the substantial quality gains driven by medical innovation. This article demonstrates how the BEA's health care satellite accounts can be combined with external data on health and life expectancy to produce quality-adjusted price indexes that better reflect changes in real output and productivity in the sector.

Consistent with a growing body of literature, our results indicate that quality-adjusted health care prices have declined relative to conventional price indexes, implying significant unmeasured productivity growth. Our estimates suggest that official statistics may understate real output growth in the health sector by roughly 1 per cent per year or more, depending on the value placed on health gains and adjustments for underlying population health trends.

These findings carry important implications. First, failing to account for health improvements can distort measures of sectoral productivity and bias assessments of living standards over time. Second, understanding how much of rising health care spending translates into improved health is central to ongoing debates over health care efficiency, pricing, and innovation policy. Improving these estimates could help better identify areas where productivity is lagging and where appropriate policy tools may be deployed. Finally, the framework presented here highlights the usefulness of condition-based health accounts and integration of health outcome data to support more accurate and actionable economic statistics.

While challenges remain, such as how to disentangle medical and non-medical contributors to health outcomes, the methodology outlined in this article provides a new, practical approach for improving the measurement of the health care sector.

References

Abraham, K.G., R.S. Jarmin, B. Moyer, and M.D. Shapiro (2021) *Big Data for Twenty-First-Century Economic Statistics*, University of Chicago Press.

Ackley, C.A. (2024) "The Impact of Preferred Provider Incentives on Demand and Negotiated Prices."

Ackley, C.A., A. Dunn, E. Dwibedi, L. Fernando, J. Joffe, J. Mallatt, C.J.L. Murray, J.L. Dieleman, and M.R. Weaver (2024) "The Value of Medical Care in the United States: Changes in Lifetime Spending and Health-Adjusted Life-Expectancy, 1996 to 2016," *Working Paper*.

Ackley, C.A., A. Dunn, and J. Romley (2026) "Quantifying Productivity Growth in Medical Care: Twenty Years of Evidence from Acute Health Conditions," Working Paper.

Ackerberg, D. and M. Rysman (2005) "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," *RAND Journal of Economics*, Vol. 36, No. 4, pp. 771–788.

Aizcorbe, A. (2006) "Why Did Semiconductor Price Indexes Fall So Fast in the 1990s? A Decomposition," *Economic Inquiry*, Vol. 44, No. 3, pp. 485–496.

Aizcorbe, A., D.M. Byrne, and D.E. Sichel (2020) "Getting Smart About Phones: New Price Indexes and the Allocation of Spending Between Devices and Services Plans in Personal Consumption Expenditures," in *Measuring Economic Growth and Productivity*, Elsevier, pp. 387–411.

Aizcorbe, A. and N. Nestoriak (2011) "Changing Mix of Medical Care Services: Stylized Facts and Implications for Price Indexes," *Journal of Health Economics*, Vol. 30, No. 3, pp. 568–574.

Aizcorbe, A. and N. Nestoriak (2012) "Price Indexes for Prescription Drugs: A Review of the Issues," *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, p. 438.

Aizcorbe, A.M. and N. Nestoriak (2010) *Price Indexes for Drugs: A Review of the Issues*, BEA.

Aizcorbe, A., G. Medeiros, and E. Strassner (2013) "Measuring Productivity for the US Health Sector," Bureau of Economic Analysis Working Paper.

Almond, D., J.J. Doyle Jr., A.E. Kowalski, and H. Williams (2010) "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns," *The Quarterly Journal of Economics*, Vol. 125, No. 2, pp. 591–634.

Anderson, G.F., U.E. Reinhardt, P.S. Hussey, and V. Petrosyan (2003) "It's the Prices, Stupid: Why the United States is So Different from Other Countries," *Health Affairs*, Vol. 22, No. 3, pp. 89–105.

Arrow, K. (1962) "Economic Welfare and the Allocation of Resources to Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton University Press, pp. 609–626.

Arrow, K.J. (1963) "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, Vol. 53, No. 5, pp. 941–973.

Baumol, W.J. (1967) "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *The American Economic Review*, Vol. 57, No. 4, pp. 415–426.

Becker, G.S., T.J. Philipson, and R.R. Soares (2005) "The Quantity and Quality of Life and the Evolution of World Inequality," *American Economic Review*, Vol. 95, No. 1, pp. 277–291.

Benner, J.S., T.W. Smith, D. Klingman, J.C. Tierce, C.D. Mullins, N. Pethick, and J.C. O'Donnell (2005) "Cost-Effectiveness of Rosuvastatin Compared with Other Statins from a Managed Care Perspective," *Value in Health*, Vol. 8, No. 6.

Berndt, E.R., D.M. Cutler, R.G. Frank, Z. Griliches, J.P. Newhouse, and J.E. Triplett (2000) "Medical Care Prices and Output," in *Handbook of Health Economics*, Vol. 1, Elsevier, pp. 119–180.

Berndt, E.R., A. Bir, S.H. Busch, R.G. Frank, and S.-L.T. Normand (2002) "The Medical Treatment of Depression, 1991–1996: Productive Inefficiency, Expected Outcome Variations, and Price Indexes," *Journal of Health Economics*, Vol. 21, No. 3, pp. 373–396.

Bhat, A.A., I. Rashid, S.U. Hassan, and P. Kansra (2023) "Social Determinants of Health and Health Outcomes: A Bibliographic Review of the Scientific Literature from 2000 to 2021," *Global Knowledge, Memory and Communication*, Vol. 74, No. 3, pp. 824–844.

Bosworth, B., J. Bieler, M. Kleinrock, E. Koepcke, and E.R. Berndt (2018) "An Evaluation of the CPI Indexes for Prescription Drugs," National Bureau of Economic Research.

Bradley, R., J. Hunjan, and L. Rozental (2015) "Experimental Disease Based Price Indexes," Bureau of Labor Statistics.

Braveman, P., S. Egerter, and D.R. Williams (2011) "The Social Determinants of Health: Coming of Age," *Annual Review of Public Health*, Vol. 32, No. 1, pp. 381–398.

Brown, C., S. Sawyer, and D. Bathgate (2020) "A Review of Hedonic Price Adjustment Techniques for Products Experiencing Rapid and Complex Quality Change," Bureau of Labor Statistics.

Bryan, K.A. and H.L. Williams (2021) "Innovation: Market Failures and Public Policies," in *Handbook of Industrial Organization*, Vol. 5, No. 1, Elsevier, pp. 281–388.

Brynjolfsson, E., A. Collis, W.E. Diewert, F. Eggers, and K.J. Fox (2019) "GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy," NBER Working Paper, No. 25695.

Byrne, D.M., S.D. Oliner, and D.E. Sichel (2018) "How Fast Are Semiconductor Prices Falling?" *Review of Income and Wealth*, Vol. 64, No. 3, pp. 679–702.

Castanheira, M., C. Ornaghi, and G. Siotis (2019) "The Unexpected Consequences of Generic Entry," *Journal of Health Economics*, Vol. 68, article 102243.

Chandra, A., J. Holmes, and J. Skinner (2013) "Is This Time Different? The Slowdown in Healthcare Spending," National Bureau of Economic Research.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016) "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector," *American Economic Review*, Vol. 106, No. 8, pp. 2110–2144.

Chandra, A. and J. Skinner (2012) "Technology Growth and Expenditure Growth in Health Care," *Journal of Economic Literature*, Vol. 50, No. 3, pp. 645–680.

Chant, E.D., M. Crawford, C.-W.W. Yang, E.S. Fisher, N.E. Morden, and I. Ganguli (2023) "Sources of Low-Value Care Received by Medicare Beneficiaries and Associated Spending Within US Health Systems," *JAMA Network Open*, Vol. 6, No. 9, pp. e2333505.

Chernew, M., A.M. Fendrick, and R.A. Hirth (1997) "Managed Care and Medical Technology: Implications for Cost Growth," *Health Affairs*, Vol. 16, No. 2, pp. 196–206.

Chernew, M.E. and J.P. Newhouse (2011) "Health Care Spending Growth," in *Handbook of Health Economics*, Vol. 2, Elsevier, pp. 1–43.

Chernew, M.E., A.L. Hicks, and S.A. Shah (2020) "Wide State-Level Variation in Commercial Health Care Prices Suggests Uneven Impact of Price Regulation," *Health Affairs*, Vol. 39, No. 5, pp. 791–799.

Chhatwal, J., F. Kanwal, M.S. Roberts, and M.A. Dunn (2015) "Cost-Effectiveness and Budget Impact of Hepatitis C Virus Treatment with Sofosbuvir and Ledipasvir in the United States," *Annals of Internal Medicine*, Vol. 162, No. 6, pp. 397–406.

Cockburn, I.M. and A.H. Anis (2001) "Hedonic Analysis of Arthritis Drugs," in *Medical Care Output and Productivity*, University of Chicago Press, pp. 439–462.

Congressional Budget Office (2022) "The Prices That Commercial Health Insurers and Medicare Pay for Hospitals' and Physicians' Services," Congressional Budget Office.

Cooper, Z., S.V. Craig, M. Gaynor, and J. Van Reenen (2019) "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured," *The Quarterly Journal of Economics*, Vol. 134, No. 1, pp. 51–107.

Curtiss, F.R. (2005) "Who Needs XR, LA, SR, XL, ER, or CR?" *Journal of Managed Care Pharmacy*, Vol. 11, No. 9.

Cutler, D.M. (1995) "Technology, Health Costs, and the NIH," Unpublished Paper Prepared for NIH Roundtable on Economics.

Cutler, D.M. (2018) "What Is the US Health Spending Problem?" *Health Affairs*, Vol. 37, No. 3, pp. 493–497.

Cutler, D.M., K. Ghosh, K.L. Messer, T.E. Raghunathan, S.T. Stewart, and A.B. Rosen (2019) "Explaining the Slowdown in Medical Spending Growth Among the Elderly, 1999–2012," *Health Affairs*, Vol. 38, No. 2, pp. 222–229.

Cutler, D.M. and D.P. Ly (2011) "The (Paper) Work of Medicine: Understanding International Medical Costs," *Journal of Economic Perspectives*, Vol. 25, No. 2, pp. 3–25.

Cutler, D.M. and M. McClellan (2001) "Is Technological Change in Medicine Worth It?" *Health Affairs*, Vol. 20, No. 5, pp. 11–29.

Cutler, D.M., M. McClellan, J.P. Newhouse, and D. Remler (1998) "Are Medical Prices Declining? Evidence from Heart Attack Treatments," *Quarterly Journal of Economics*, Vol. 113, No. 4, pp. 991–1024.

Cutler, D.M., A.B. Rosen, and S. Vijan (2006) "The Value of Medical Spending in the United States, 1960–2000," *New England Journal of Medicine*, Vol. 355, No. 9, pp. 920–927.

Cutler, D.M., K. Ghosh, K.L. Messer, T. Raghunathan, A.B. Rosen, and S.T. Stewart (2022) "A Satellite Account for Health in the United States," *American Economic Review*, Vol. 112, No. 2, pp. 494–533.

Dauda, S., A. Dunn, and A. Hall (2022) "A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care Using Claims Data," *Journal of Health Economics*, Vol. 85, article 102662.

Dattani, S., L. Rodés-Guirao, H. Ritchie, E. Ortiz-Ospina, and M. Roser (2023) "Life Expectancy," *Our World in Data*.

Deaton, A. and P. Schreyer (2022) "GDP, Wellbeing, and Health: Thoughts on the 2017 Round of the International Comparison Program," *Review of Income and Wealth*, Vol. 68, No. 1, pp. 1–15.

Dieleman, J.L., J. Cao, A. Chapin, C. Chen, Z. Li, A. Liu, C. Horst, A. Kaldjian, T. Matyasz, K.W. Scott, and others (2020) "US Health Care Spending by Payer and Health Condition, 1996–2016," *JAMA*, Vol. 323, No. 9, pp. 863–884.

Dieleman, J.L., M. Beauchamp, S.W. Crosby, D. DeJarnatt, E.K. Johnson, H. Lescinsky, T. McHugh, I. Pollock, M. Sahu, V. Swart, and others (2025) "Tracking US Health Care Spending by Health Condition and County," *JAMA*.

Dilokthornsakul, P., M. Patidar, and J.D. Campbell (2017) "Forecasting the Long-Term Clinical and Economic Outcomes of Lumacaftor/Ivacaftor in Cystic Fibrosis Patients with Homozygous Phe508del Mutation," *Value in Health*, Vol. 20, No. 10, pp. 1329–1335.

DiMasi, J.A., R.W. Hansen, and H.G. Grabowski (2003) "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics*, Vol. 22, No. 2, pp. 151–185.

DiMasi, J.A., H.G. Grabowski, and R.W. Hansen (2016) "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs," *Journal of Health Economics*, Vol. 47, pp. 20–33.

Duan, N. (1983) "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, Vol. 78, No. 383, pp. 605–610.

Duflos, G. and F.R. Lichtenberg (2012) "Does Competition Stimulate Drug Utilization? The Impact of Changes in Market Structure on US Drug Prices, Marketing and Utilization," *International Review of Law and Economics*, Vol. 32, No. 1, pp. 95–109.

Dunn, A., L. Fernando, and E. Liebman (2023) "A Direct Measure of Medical Innovation on Health Care Spending: A Condition-Specific Approach," *Health Management Policy and Innovation*, Vol. 8, No. 2.

Dunn, A., L. Fernando, and E. Liebman (2024) "How Much Are Medical Innovations Worth? A Detailed Analysis Using Thousands of Cost-Effectiveness Studies."

Dunn, A., A. Hall, and S. Dauda (2022) "Are Medical Care Prices Still Declining? A Re-Examination Based on Cost-Effectiveness Studies," *Econometrica*, Vol. 90, No. 2, pp. 859–886.

Dunn, A., L. Rittmueller, and B. Whitmire (2015) “Introducing the New BEA Health Care Satellite Account,” *Survey of Current Business*, Vol. 95, No. 1, pp. 1–21.

Dunn, A., L. Rittmueller, and B. Whitmire (2016) “Health Care Spending Slowdown from 2000 to 2010 Was Driven by Lower Growth in Cost per Case, According to a New Data Source,” *Health Affairs*, Vol. 35, No. 1, pp. 132–140.

Dunn, A., B. Whitmire, A. Batch, L. Fernando, and L. Rittmueller (2018) “High Spending Growth Rates for Key Diseases in 2000–14 Were Driven by Technology and Demographic Factors,” *Health Affairs*, Vol. 37, No. 6, pp. 915–924.

Dynan, K. and L. Sheiner (2018) “GDP as a Measure of Economic Well-Being,” Hutchins Center Working Paper, No. 43.

Eggleston, K., B.K. Chen, C.-H. Chen, Y.I. Chen, T. Feenstra, T. Iizuka, J.T. Lam, G.M. Leung, J.-f.R. Lu, B. Rodriguez-Sanchez, and others (2020) “Are Quality-Adjusted Medical Prices Declining for Chronic Disease? Evidence from Diabetes Care in Four Health Systems,” *The European Journal of Health Economics*, Vol. 21, pp. 689–702.

Einav, L. and A. Finkelstein (2023) *We’ve Got You Covered: Rebooting American Health Care*, Penguin.

Feenstra, R.C. (1994) “New Product Varieties and the Measurement of International Prices,” *The American Economic Review*, pp. 157–177.

Fiore, J.A., A.J. Madison, J.A. Poisal, G.A. Cuckler, S.D. Smith, A.M. Sisko, S.P. Keehan, K.E. Rennie, and A.C. Gross (2024) “National Health Expenditure Projections, 2023–32: Payer Trends Diverge as Pandemic-Related Policies Fade,” *Health Affairs*, Vol. 43, No. 7, pp. 910–921.

Finkelstein, A., M. Gentzkow, and H. Williams (2016) “Sources of Geographic Variation in Health Care: Evidence from Patient Migration,” *The Quarterly Journal of Economics*, Vol. 131, No. 4, pp. 1681–1726.

Fisher, E.S., D.E. Wennberg, T.A. Stukel, D.J. Gottlieb, F.L. Lucas, and E.L. Pinder (2003a) “The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care,” *Annals of Internal Medicine*, Vol. 138, No. 4, pp. 273–287.

Fisher, E.S., D.E. Wennberg, T.A. Stukel, D.J. Gottlieb, F.L. Lucas, and E.L. Pinder (2003b) “The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care,” *Annals of Internal Medicine*, Vol. 138, No. 4, pp. 288–298.

Fisher, F.M. and K. Shell (1972) *The Economic Theory of Price Indices*, Academic Press, New York and London.

Frank, R.G., E.R. Berndt, A.B. Busch, and A.F. Lehman (2004) “Quality-Constant ‘Prices’ for the Ongoing Treatment of Schizophrenia: An Exploratory Study,” *The Quarterly Review of Economics and Finance*, Vol. 44, No. 3, pp. 390–409.

Garber, A.M. and J. Skinner (2008) “Is American Health Care Uniquely Inefficient?” *Journal of Economic Perspectives*, Vol. 22, No. 4, pp. 27–50.

Gastala, N.M., P. Wingrove, A. Gaglioti, S. Petterson, and A. Bazemore (2016) “Medicare Part D: Patients Bear the Cost of ‘Me Too’ Brand-Name Drugs,” *Health Affairs*, Vol. 35, No. 7.

Giugliano, R.P., T.R. Pedersen, J.L. Saver, P.S. Sever, A.C. Keech, E.A. Bohula, S.A. Murphy, S.M. Wasserman, N. Honarpour, H. Wang, A.L. Pineda, and M.S. Sabatine (2020) “Stroke Prevention with the PCSK9 Inhibitor Evolocumab Added to Statin in High-Risk Patients with Stable Atherosclerosis,” *Stroke*, Vol. 51, No. 5.

Glied, S. and A. Sacarny (2018) “Is the US Health Care System Wasteful and Inefficient? A Review of the Evidence,” *Journal of Health Politics, Policy and Law*, Vol. 43, No. 5, pp. 739–765.

Goodman, D.C., E.S. Fisher, J.E. Wennberg, and J.S. Skinner (2008) “Tracking the Care of Patients with Severe Chronic Illness: The Dartmouth Atlas of Health Care 2008.”

Gowrisankaran, G. and M. Rysman (2012) “Dynamics of Consumer Demand for New Durable Goods,” *Journal of Political Economy*, Vol. 120, No. 6, pp. 1173–1219.

Groshen, E.L., B.C. Moyer, A.M. Aizcorbe, R. Bradley, and D.M. Friedman (2017) “How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches,” *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 187–210.

Grossman, M. (1972) “On the Concept of Health Capital and the Demand for Health,” *Journal of Political Economy*, Vol. 80, No. 2, pp. 223–255.

Hall, A.E. (2017) “Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature,” *Medical Care Research and Review*, Vol. 74, No. 6, pp. 639–667.

Hall, R.E. and C.I. Jones (2007) “The Value of Life and the Rise in Health Spending,” *The Quarterly Journal of Economics*, Vol. 122, No. 1, pp. 39–72.

Hartman, M., A.B. Martin, B. Washington, A. Catlin, and the National Health Expenditure Accounts Team (2021a) “National Health Care Spending in 2020: Growth Driven by Federal Spending in Response to the COVID-19 Pandemic,” *Health Affairs*, Vol. 41, No. 9.

Hartman, M., A.B. Martin, D. Lassman, A. Catlin, and the National Health Expenditure Accounts Team (2021b) "National Health Care Spending in 2019: Steady Growth for the Fourth Consecutive Year," *Health Affairs*, Vol. 40, No. 1, pp. 14–24.

Hemphill, C.S. and B.N. Sampat (2012) "Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals," *Journal of Health Economics*, Vol. 31, No. 2.

Highfill, T. and E. Bernstein (2019) "Using Disability Adjusted Life Years to Value the Treatment of Thirty Chronic Conditions in the US from 1987 to 2010: A Proof of Concept," *International Journal of Health Economics and Management*, Vol. 19, No. 3–4, pp. 449–466.

Ho, K. and A. Pakes (2024) "Evaluating Pharmaceutical Policy Options," National Bureau of Economic Research.

Hofmeister, M.G., E.M. Rosenthal, L.K. Barker, E.S. Rosenberg, M.A. Barranco, E.W. Hall, B.R. Edlin, J. Mermin, J.W. Ward, and A.B. Ryerson (2019) "Estimating Prevalence of Hepatitis C Virus Infection in the United States, 2013–2016," *Hepatology*, Vol. 69, No. 3, pp. 1020–1031.

Hopkins, J. (2021) "House Committee Takes On Pricing, Patents for Top-Selling Drug Humira," *Wall Street Journal*.

Hult, K.J., S. Jaffe, and T.J. Philipson (2018) "How Does Technological Change Affect Quality-Adjusted Prices in Health Care? Systematic Evidence from Thousands of Innovations," *American Journal of Health Economics*, Vol. 4, No. 4, pp. 433–453.

January, C.T., J.S. Alpert, H. Calkins, J.E. Cigarroa, J.C. Cleveland, J.B. Conti, P.T. Ellinor, M.D. Ezekowitz, M.E. Field, K.T. Murray, R.L. Sacco, W.G. Stevenson, C.M. Tracy, and C.W. Yancy (2014) "2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation," *Circulation*, Vol. 130, No. 23.

January, C.T. and S.L. Wann (2019) "2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation," *Circulation*, Vol. 140, No. 2.

Jaruzelski, B., R. Chwalik, and B. Goehle (2018) "What the Top Innovators Get Right," *strategy+ business*, Vol. 93.

Jena, A.B. and T.J. Philipson (2008) "Cost-Effectiveness Analysis and Innovation," *Journal of Health Economics*, Vol. 27, No. 5, pp. 1224–1236.

Kakani, P., M. Chernew, and A. Chandra (2020) "Rebates in the Pharmaceutical Industry: Evidence from Medicines Sold in Retail Pharmacies in the U.S.," NBER Working Paper, No. 26846.

Kaplan, R.M. and A. Milstein (2019) "Contributions of Health Care to Longevity: A Review of 4 Estimation Methods," *The Annals of Family Medicine*, Vol. 17, No. 3, pp. 267–272.

Kearsley, A. (2024) "HHS Standard Values for Regulatory Analysis, 2024," Technical Report, Assistant Secretary for Planning and Evaluation, Health and Human Services.

Kesselheim, A.S., J. Avorn, and A. Sarpatwari (2016) "The High Cost of Prescription Drugs in the United States: Origins and Prospects for Reform," *JAMA*, Vol. 316, No. 8, pp. 858–871.

Kyle, M. and H. Williams (2017) "Is American Health Care Uniquely Inefficient? Evidence from Prescription Drugs," *American Economic Review*, Vol. 107, No. 5, pp. 486–490.

Lakdawalla, D., A. Malani, and J. Reif (2017) "The Insurance Value of Medical Innovation," *Journal of Public Economics*, Vol. 145, pp. 94–102.

Lebow, D.E. and J.B. Rudd (2003) "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature*, Vol. 41, No. 1, pp. 159–201.

Lichtenberg, F.R. (2024) "Has Pharmaceutical Innovation Reduced the Average Cost of US Health Care Episodes?" *International Journal of Health Economics and Management*, Vol. 24, No. 1, pp. 1–31.

Lin, G., M. Whittington, D. Niktin, F. Agboola, A. McKenna, S. Herron-Smith, S. Pearson, and J. Campbell (2023) "Oral and Monoclonal Antibody Treatments for Relapsing Forms of Multiple Sclerosis: Effectiveness and Value," Institute for Clinical and Economic Review.

Lopez, C., H. Roh, and B. Butler (2020) "How to Identify Health Innovation Gaps?"

Lu, Z.J. and W.S. Comanor (1998) "Strategic Pricing of New Pharmaceuticals," *Review of Economics and Statistics*, Vol. 80, No. 1, pp. 108–118.

Lucarelli, C. and S. Nicholson (2009) "A Quality-Adjusted Price Index for Colorectal Cancer Drugs," National Bureau of Economic Research.

Lucarelli, C., S. Nicholson, and N. Tilipman (2022) "Price Indices and the Value of Innovation with Heterogeneous Patients," *Journal of Health Economics*, Vol. 84, article 102625.

Mahipal, A. and A. Grothey (2016) "Role of Biologics in First-Line Treatment of Colorectal Cancer," *Journal of Oncology Practice*, Vol. 12, No. 12.

Mallatt, J., A. Dunn, and L. Fernando (2024) "Consumer Out-Of-Pocket Drug Prices Grew Faster Than Prices Faced By Insurers After Accounting For Rebates, 2007–20," *Health Affairs*, Vol. 43, No. 9, pp. 1284–1289.

Marmot, M. and R.G. Wilkinson (2006) "Social Determinants of Health, 2nd Edition," *International Journal of Epidemiology*, Vol. 35, No. 4, p. 1111.

Matsumoto, B. and others (2021) *Producing Quality Adjusted Hospital Price Indexes*, US Department of Labor, US Bureau of Labor Statistics.

Meltzer, D. (1997) "Accounting for Future Costs in Medical Cost-Effectiveness Analysis," *Journal of Health Economics*, Vol. 16, No. 1, pp. 33–64.

Meltzer, D.O. and P.C. Smith (2011) "Theoretical Issues Relevant to the Economic Evaluation of Health Technologies," in *Handbook of Health Economics*, Vol. 2, Elsevier, pp. 433–469.

Millman, J. (2014a) "This Drug Costs $1,125 per Pill and Is About to Shatter Sales Records," *The Washington Post*.

Millman, J. (2014b) "Does It Really Cost $2.6 Billion to Develop a New Drug?"

Milken Institute (2020) "How to Identify Health Innovation Gaps? Insights from Data on Diseases' Costs, Mortality, and Funding," Working Paper.

Mokdad, A.H., C. Bisignano, J.M. Hsu, H.S. Ababneh, R. Abbasgholizadeh, A. Abdelkader, and others (2024) "The Burden of Diseases, Injuries, and Risk Factors by State in the USA, 1990–2021: A Systematic Analysis for the Global Burden of Disease Study 2021," *The Lancet*, Vol. 404, No. 10469, pp. 2314–2340.

Montero, A., A. Kearney, L. Hamel, and M. Brodie (2022) "Americans' Challenges with Health Care Costs," Kaiser Family Foundation.

Murphy, K.M. and R.H. Topel (2006) "The Value of Health and Longevity," *Journal of Political Economy*, Vol. 114, No. 5, pp. 871–904.

National Research Council (2002) *At What Price?: Conceptualizing and Measuring Cost-of-Living and Price Indexes*, National Academies Press.

National Research Council (2011) *Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement*, National Academies Press.

Nelson, R.R. (1959) "The Simple Economics of Basic Scientific Research," *Journal of Political Economy*, Vol. 67, No. 3, pp. 297–306.

Neumann, P.J., J.T. Cohen, and M.C. Weinstein (2014) "Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold," *New England Journal of Medicine*, Vol. 371, No. 9, pp. 796–797.

Newhouse, J.P. (1992) "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives*, Vol. 6, No. 3, pp. 3–21.

Nolte, E. and M. McKee (2011) "Variations in Amenable Mortality — Trends in 16 High-Income Nations," *Health Policy*, Vol. 103, No. 1, pp. 47–52.

Nordhaus, W.D. (1972) "Is Growth Obsolete?" *Economic Research*.

Nordhaus, W.D. (2004) "Schumpeterian Profits in the American Economy: Theory and Measurement," National Bureau of Economic Research.

Olek, M.J. and E. Mowry (2022) "Initial Disease-Modifying Therapy for Relapsing-Remitting Multiple Sclerosis in Adults," UpToDate.

Ortiz-Ospina, E. and M. Roser (2017) "Healthcare Spending," *Our World in Data*.

Pakes, A. (2003) "A Reconsideration of Hedonic Price Indexes with an Application to PC's," *American Economic Review*, Vol. 93, No. 5, pp. 1578–1596.

Park, S., J.L. Dieleman, M.R. Weaver, G. Bae, and K. Eggleston (2025) "Health Care Spending Increases and Value in South Korea," *JAMA Health Forum*, Vol. 6, No. 1, pp. e245145.

Peter G. Peterson Foundation (2020) "How Does the US Healthcare System Compare to Other Countries?"

Petrin, A. (2002) "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, Vol. 110, No. 4, pp. 705–729.

Philipson, T.J. and A.B. Jena (2006) "Who Benefits from New Medical Technologies? Estimates of Consumer and Producer Surpluses for HIV/AIDS Drugs," *Forum for Health Economics & Policy*, Vol. 9, No. 2.

Preston, S.H. and A. Stokes (2011) "Contribution of Obesity to International Differences in Life Expectancy," *American Journal of Public Health*, Vol. 101, No. 11, pp. 2137–2143.

Ridker, P.M., E. Danielson, F.A.H. Fonseca, J. Genest, A.M. Gotto Jr., J.J.P. Kastelein, W. Koenig, P. Libby, A.J. Lorenzatti, J.G. MacFadyen, B.G. Nordestgaard, J. Shepherd, J.T. Willerson, and R.J. Glynn (2008) "Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated C-Reactive Protein," *New England Journal of Medicine*, Vol. 359, No. 21.

Ridley, D.B. and S. Zhang (2017) "Regulation of Price Increases," *International Journal of Industrial Organization*, Vol. 50, pp. 186–213.

Rind, D., D.A. Ollendorf, R. Chapman, and others (2016) "Treatment Options for Advanced Non-Small Cell Lung Cancer: Effectiveness, Value and Value-Based Price Benchmarks," Institute for Clinical and Economic Review.

Roehrig, C. and M. Daly (2015) "Prevalence Trends for Three Common Medical Conditions: Treated and Untreated," *Health Affairs*, Vol. 34, No. 8, pp. 1320–1323.

Romley, J.A., A. Dunn, D. Goldman, and N. Sood (2020) "Quantifying Productivity Growth in the Delivery of Important Episodes of Care Within the Medicare Program Using Insurance Claims and Administrative Data," in *Big Data for Twenty-First-Century Economic Statistics*, University of Chicago Press.

Romley, J.A., D.P. Goldman, and N. Sood (2015) "US Hospitals Experienced Substantial Productivity Growth During 2002–11," *Health Affairs*, Vol. 34, No. 3, pp. 511–518.

Sanger-Katz, M. (2021) "U.S. Health Spending Reaches Record High in 2020," *The New York Times*.

Schmalensee, R. (1982) "Product Differentiation Advantages of Pioneering Brands," *The American Economic Review*, Vol. 72, No. 3, pp. 349–365.

Schreyer, P. (2012) "Output, Outcome, and Quality Adjustment in Measuring Health and Education Services," *Review of Income and Wealth*, Vol. 58, No. 2, pp. 257–278.

Schroeder, S.A. (2007) "We Can Do Better — Improving the Health of the American People," *New England Journal of Medicine*, Vol. 357, No. 12, pp. 1221–1228.

Schwartz, W.B. (1987) "The Inevitable Failure of Current Cost-Containment Strategies: Why They Can Provide Only Temporary Relief," *JAMA*, Vol. 257, No. 2, pp. 220–224.

Shapiro, I., M.D. Shapiro, and D.W. Wilcox (2001) "A Price Index for Cataract Surgery," in *Medical Output and Productivity*, University of Chicago Press for the National Bureau of Economic Research, pp. 411–438.

Sharpe, A., H. Messinger, and C. Bradley (2007) *The Measurement of Output and Productivity in the Health Care Sector in Canada: An Overview*, Centre for the Study of Living Standards.

Sharma, D., S. Xing, Y.-T. Hung, R.N. Caskey, M.L. Dowell, and D.R. Touchette (2018) "Cost-Effectiveness Analysis of Lumacaftor and Ivacaftor Combination for the Treatment of Patients with Cystic Fibrosis in the United States," *Orphanet Journal of Rare Diseases*, Vol. 13, pp. 1–12.

Sheiham, A. (2009) "Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health," *Community Dental Health*, Vol. 26, No. 1, pp. 2–3.

Sheiner, L. (2014) "Why the Geographic Variation in Health Care Spending Cannot Tell Us Much About the Efficiency or Quality of Our Health Care System," *Brookings Papers on Economic Activity*, Vol. 2014, No. 2, pp. 1–72.

Sheiner, L. and D.M. Cutler (2024) "Measuring Prices and Productivity in the Health Care Sector," in *The Measure of Economies: Measuring Productivity in an Age of Technological Change*, University of Chicago Press, pp. 225–262.

Sheiner, L. and A. Malinovskaya (2016) "Measuring Productivity in Healthcare: An Analysis of the Literature," Hutchins Center on Fiscal and Monetary Policy at Brookings.

Shrank, W.H., T.L. Rogstad, and N. Parekh (2019) "Waste in the US Health Care System: Estimated Costs and Potential for Savings," *JAMA*, Vol. 322, No. 15, pp. 1501–1509.

Small, K.A. and H.S. Rosen (1981) "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, pp. 105–130.

Smith, S., J.P. Newhouse, and M.S. Freeland (2009) "Income, Insurance, and Technology: Why Does Health Spending Outpace Economic Growth?" *Health Affairs*, Vol. 28, No. 5, pp. 1276–1284.

Smith, S.D., J.P. Newhouse, and G.A. Cuckler (2022) "Health Care Spending Growth Has Slowed: Will the Bend in the Curve Continue?" NBER Working Paper, No. 30782.

Solow, R.M. (1957) "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, pp. 312–320.

Syverson, C. (2017) "Challenges to Mismeasurement Explanations for the US Productivity Slowdown," *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 165–186.

Tabernero, J. (2015) "Proven Efficacy, Equitable Access, and Adjusted Pricing of Anti-Cancer Therapies: No 'Sweetheart' Solution," *Annals of Oncology*, Vol. 26, No. 8.

The Centers for Medicare and Medicaid Services (2015) "Medicare Drug Spending Dashboard," Fact Sheet.

The Centers for Medicare and Medicaid Services (2021) "Medicare Part D Spending by Drug (Historical); Medicaid Spending by Drug (Historical)."

Tice, J.A., D.A. Ollendorf, H.S. Chahal, J.G. Kahn, E. Marseille, J. Weissberg, K.K. Shore, and S.D. Pearson (2015) "The Comparative Clinical Effectiveness and Value of Novel Combination Therapies for the Treatment of Patients with Genotype 1 Chronic Hepatitis C Infection: A Technology Assessment," Institute for Clinical and Economic Review.

Tice, J.A., R. Chapman, V. Kumar, P. Synnott, M. Seidner, and D.A. Ollendorf (2017) "Anabolic Therapies for Osteoporosis in Postmenopausal Women: Effectiveness and Value," Institute for Clinical and Economic Review.

Tice, J.A., R. Chapman, M. Seidner, D.M. Rind, S.D. Pearson, K.M. Kuntz, and K. Wherry (2020) "Modulator Treatments for Cystic Fibrosis: Effectiveness and Value," Institute for Clinical and Economic Review.

Tice, J.A., M. Richardson, A. Wright, M. Seidner, D.M. Rind, and S.D. Pearson (2023) "Special Assessment to Inform CMS Drug Price Negotiation: Eliquis and Xarelto," Institute for Clinical and Economic Review.

Tito, F., S. Mailankody, and A. Lo (2014) "Unintended Consequences of Expensive Cancer Therapeutics — The Pursuit of Marginal Indications and a Me-Too Mentality That Stifles Innovation and Creativity," *JAMA Otolaryngology — Head & Neck Surgery*, Vol. 140, No. 12.

Trajtenberg, M. (1989) "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, Vol. 97, No. 2, pp. 444–479.

Trajtenberg, M. (1990) "Product Innovations, Price Indices and the (Mis)measurement of Economic Performance," National Bureau of Economic Research.

Trogdon, J.G., E.A. Finkelstein, and T.J. Hoerger (2008) "Use of Econometric Models to Estimate Expenditure Shares," *Health Services Research*, Vol. 43, No. 4, pp. 1442–1452.

van der Gronde, T., C.A. Uyl-de Groot, and T. Pieters (2017) "Addressing the Challenge of High-Priced Prescription Drugs in the Era of Precision Medicine: A Systematic Review of Drug Life Cycles, Therapeutic Drug Markets and Regulatory Frameworks," *PLoS ONE*, Vol. 12, No. 8.

Viscusi, W.K. (2020) "Failing to Think Properly About the Value of a Statistical Life," *The Regulatory Review*.

Vos, T., S.S. Lim, C. Abbafati, K.M. Abbas, M. Abbasi, M. Abbasifard, M. Abbasi-Kangevari, H. Abbastabar, F. Abd-Allah, A. Abdelalim, and others (2020) "Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019," *The Lancet*, Vol. 396, No. 10258, pp. 1204–1222.

Weaver, M.R., J. Joffe, M. Ciarametaro, R.W. Dubois, A. Dunn, A. Singh, G.W. Sparks, L. Stafford, C.J.L. Murray, and J.L. Dieleman (2022) "Health Care Spending Effectiveness: Estimates Suggest That Spending Improved US Health From 1996 to 2016," *Health Affairs*, Vol. 41, No. 7, pp. 994–1004.

Wennberg, J.E., M.M. Cooper, T.A. Bubolz, E.S. Fisher, A.M. Gittelsohn, D.C. Goodman, J.E. Mohr, J.F. Poage, S.M. Sharp, J. Skinner, and T.A. Stukel (1996) *The Dartmouth Atlas of Health Care in the United States 1996*, American Hospital Publishing, Inc., Chicago, IL.

West Health and Gallup (2021) *2021 Health Care in America Report*.

Addressing Canada's High Cost of Living: The Role of Productivity and Bargaining Power

Claude Lavoie*
IPM Commentary

Abstract

This commentary examines persistent concerns about the cost of living among Canadians, a phenomenon also observed across many developed economies. Despite macroeconomic indicators showing that household income has generally grown faster than prices in recent years, approximately 60 per cent of Canadians identify the cost of living as a primary concern. Drawing on polling data, economic statistics, and policy literature, the analysis identifies housing affordability, slowing real income growth, and social media–driven financial perceptions as key drivers. The commentary concludes that policy responses should prioritize housing supply, productivity growth, and stronger worker bargaining power to ensure that economic gains translate into improved household welfare.

1. Introduction

The cost of living has emerged as the predominant concern for Canadian citizens, with 62 per cent identifying it as a top issue in October 2025, according to Abacus polling data. This represents a consistent pattern, as similar polling in 2019 and in 2015 showed 55 per cent to 60 per cent of Canadians expressing the same concern (Table 1). The persistence of this anxiety, both before and after the COVID-19 pandemic, suggests structural rather than cyclical causes.

This concern is not unique to Canada. Approximately 45 per cent of French citizens identify cost of living as a primary issue (also in line with the pre-pandemic period), while similar patterns appear across the United States and other European nations. The international scope of this phenomenon suggests that broader economic forces, rather than country-specific policies, are at play.

This commentary examines the sources of widespread concern about cost of living. It is unclear how people, when surveyed, define the "cost of living" and what specifically concerns them. This commentary assumes, consistent with the literature, that such concerns arise when households feel that the prices of day-to-day consumption

* Claude Lavoie is a Former Director General in the Economic and Fiscal Policy Branch at Finance Canada. Email: claudeblavoie@gmail.com

Table 1: Top Issues Facing Canada

Per cent of Canadian survey respondents

Issue	July 2015	October 2019	October 2025
Cost of living	55	60	62
Health care	N/A	44	35
The economy	36	29	35
Housing affordability	N/A	27	34
Donald Trump	N/A	N/A	33
Immigration	N/A	26	27

Note: Because respondents identify multiple issues, the percentages in any given survey will sum to more than 100.

Source: Abacus Polls for 2025 and 2019 and Ipsos for 2015. Percentage of responses to the question: "What are the 3 most important issues facing Canada today?" • Created with Datawrapper

are rising faster than their current (or expected) income. This commentary seeks to find the potential sources of these concerns, rather than determine whether a true cost of living crisis exists.

2. Potential Causal Factors

2.1 Inflation: A Limited Explanation

When citizens express concern about cost of living, they primarily reference the prices of everyday goods and services. Theoretically, rapid price increases — high inflation — should correlate with such concerns. However, Canada's inflation performance does not support this hypothesis.

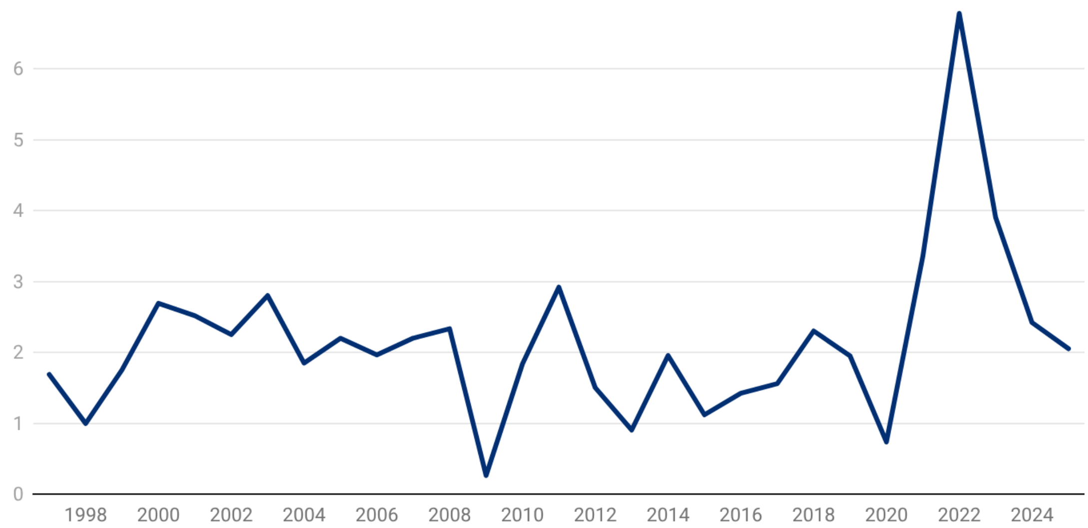
Canada has maintained inflation close to its 2 per cent target, averaging 2.1 per cent since 1995 and 2.3 per cent since the 2009 financial crisis. While pandemic-era inflation exceeded 3 per cent annually, pushing prices approximately 6 per cent above target trajectory, cost of living concerns preceded this period (Figure 1). In 2019, when inflation measured only 1.7 per cent, 60 per cent of Canadians already identified cost of living as their primary concern. Furthermore, these concerns remain even though inflation expectations have since normalized, with consumers anticipating inflation rates near the 2 per cent target going forward. These patterns suggest that general inflation cannot explain the persistent and widespread anxiety about the cost of living.

2.2 Housing Affordability: A Significant Contributor

Certain price categories receive disproportionately more attention in household budgets, particularly necessities such as food, shelter, and transportation. As Schembri (2020) notes, consumers' perceptions of inflation may be higher than officially measured inflation rates if the prices of frequently-purchased goods are rising faster than generalized inflation, or if they put more weight on some goods and services whose prices are rising. Higher

Figure 1: Inflation Rate, 1997-2025

Per cent annual change in the price level



Source: Statistics Canada, Table: 18-10-0005-01 • Created with Datawrapper

inflation for these products may raise an alarm about the broader affordability of basic necessities. Analysis of these categories reveals differential patterns (Table 2).

Food and gasoline prices accelerated post-pandemic but remained relatively modest during the pre-pandemic period when such concerns already existed. Housing presents a more complex picture. Consumers' perceptions of inflation are likely influenced by the prices of houses, while the CPI measures the prices of housing — more specifically, the costs associated with home ownership, which includes rent and mortgage payments but not directly the price of a house.

The costs of renting or owning a house have followed a similar pattern to those of food and gasoline but house purchase prices increased approximately 5 per cent annually over the past decade, both pre- and post-pandemic. Metropolitan rental markets — particularly Toronto, Vancouver, and Ottawa — also experienced similar growth rates despite more moderate national trends.

However, the two-thirds of Canadians who own homes generally benefit from house price appreciation through increased wealth. In principle, this should make them feel they have higher purchasing power, all else equal, rather than reduce affordability. The remaining third, comprising renters in high-cost markets, prospective first-time buyers, and those who over-extended financially during purchase, experience genuine affordability pressure. This segment broadly corresponds to the 34 per cent identifying housing affordability as a top concern (Table 1), suggesting housing constitutes a significant but not comprehensive explanation for broader cost of living anxiety.

2.3 The Income-Perception Paradox

As noted earlier, cost of living concerns arise when households feel that the prices

Table 2: Growth in Selected Prices in Canada

Annual per cent change

Category	2010-19	2019-24	2010-24
All Items	1.7	3.4	2.3
Food	2.2	4.7	3.1
Gasoline	1.6	6.0	3.1
Shelter	1.8	4.8	2.8

Source: Statistics Canada, Table: 18-10-0005-01 • Created with Datawrapper

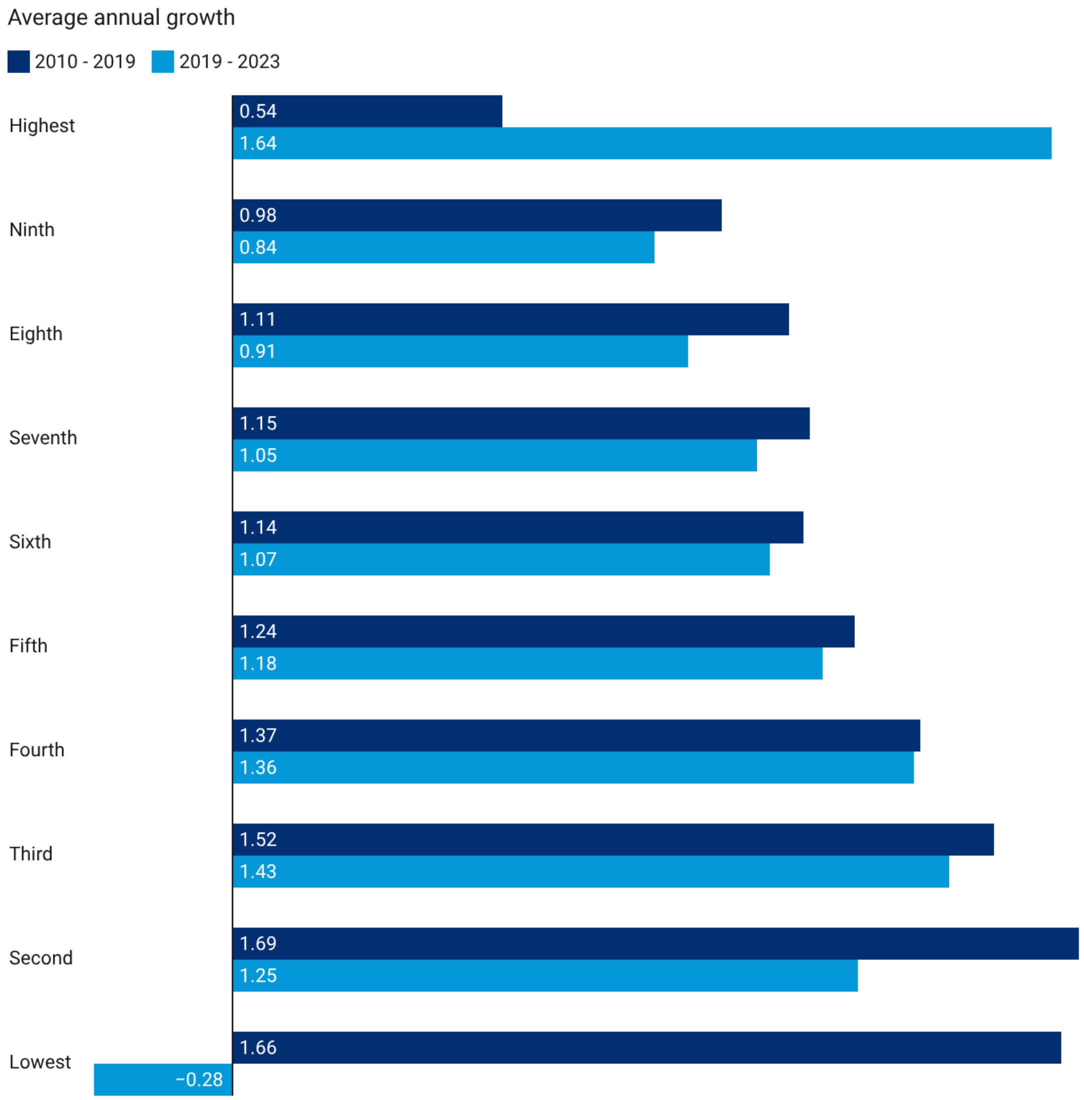
of day-to-day consumption items are rising faster than their income, in other words when their real income is perceived to be declining. Yet, most Canadians experienced real income growth, despite this widespread perception of the contrary (Figure 2).

Real disposable income for households in every income decile — except the lowest — was higher in 2023 than in 2019 (pre-pandemic) and substantially higher than in 2010 (post-recession). The temporary elevation of pandemic-era incomes, driven by exceptional government transfers, cannot explain current sentiment, as cost of living concerns preceded COVID-19, and no reasonable expectation existed for permanent continuation of emergency support.

Households do not directly observe disposable income without some calculation, which may cause perceptions to diverge from reality. Assessing disposable income requires aggregating multiple components, including earnings, taxes, and transfers. People may also suffer from money illusion and, as Krugman (2025) notes, may see their pay increases, not as reflecting higher inflation, but instead as reflecting their hard work. Still, people do observe the wage they receive and the prices they pay. Real hourly market wages increased across most occupations during both 2019-2025 and 2010-2025 periods. Declines occurred in certain occupations, but these were modest and affected only 10-15 per cent of the labour force. Additionally, occupational mobility suggests that some workers in affected sectors may have transitioned to alternative employment, partly mitigating individual wage declines.

Aggregate data do not appear to conceal significant population groups experiencing acute cost-of-living pressures. Real income growth of younger cohorts, particularly those aged 25-34, lagged behind older citizens but did not experience absolute decline. Young families, whose head of the household is aged between 25 and 34, and who represent approximately 15 per cent of the population, saw median real income remain flat between 2019 and 2023 — essentially with no gains, but also no losses for this group overall. Moreover, the proportion of households experiencing significant income decline (exceeding 10 per cent over five-year periods) remained stable at

Figure 2: Real Household Disposable Income per Decile



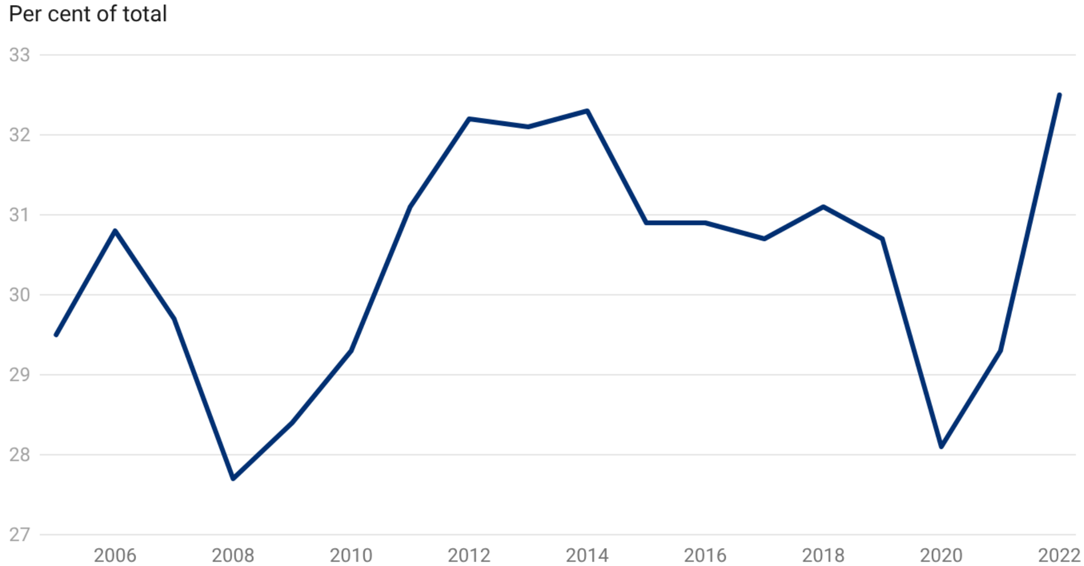
Source: Statistics Canada: Table 11-10-0193-01 • Created with Datawrapper

approximately 30 per cent since the mid-2000s (Figure 3).

Cost of living rising faster than incomes should be expected to deplete Canadians' financial wealth. A January 2025 RBC poll (Royal Bank of Canada, 2025) found that over half of Canadians describe themselves as "financially paralyzed."

But net worth statistics contradict these narratives. Over roughly the past decade, households across all income deciles experienced significant real net worth increases (Figure 4). This growth reflects not only housing and pension asset appreciation, but also increases in non-pension financial assets net of non-mortgage debt, as net financial worth, which excludes them, also improves across all deciles. International evidence suggests similar patterns exist in other developed nations.

Figure 3: Tax Filers With Lower After-Tax Income Than 5 Years Earlier



Source: Statistics Canada, Table: 11-10-0059-01 • Created with Datawrapper

2.4 The Growth Deceleration Hypothesis

While income grew faster than prices for most Canadians, the pace of improvement decelerated markedly (Figure 5). Households in most income deciles experienced real disposable income growth approximately 50 per cent slower during 2009-2023 compared to 1995-2009. Real hourly wages, median incomes, and age-cohort income data corroborate this deceleration pattern.

This phenomenon extends beyond Canada. International Monetary Fund Managing Director Christine Lagarde characterized the post-2010 period as "the new mediocre," acknowledging widespread productivity and growth slowdowns across developed economies (Lagarde, 2014).

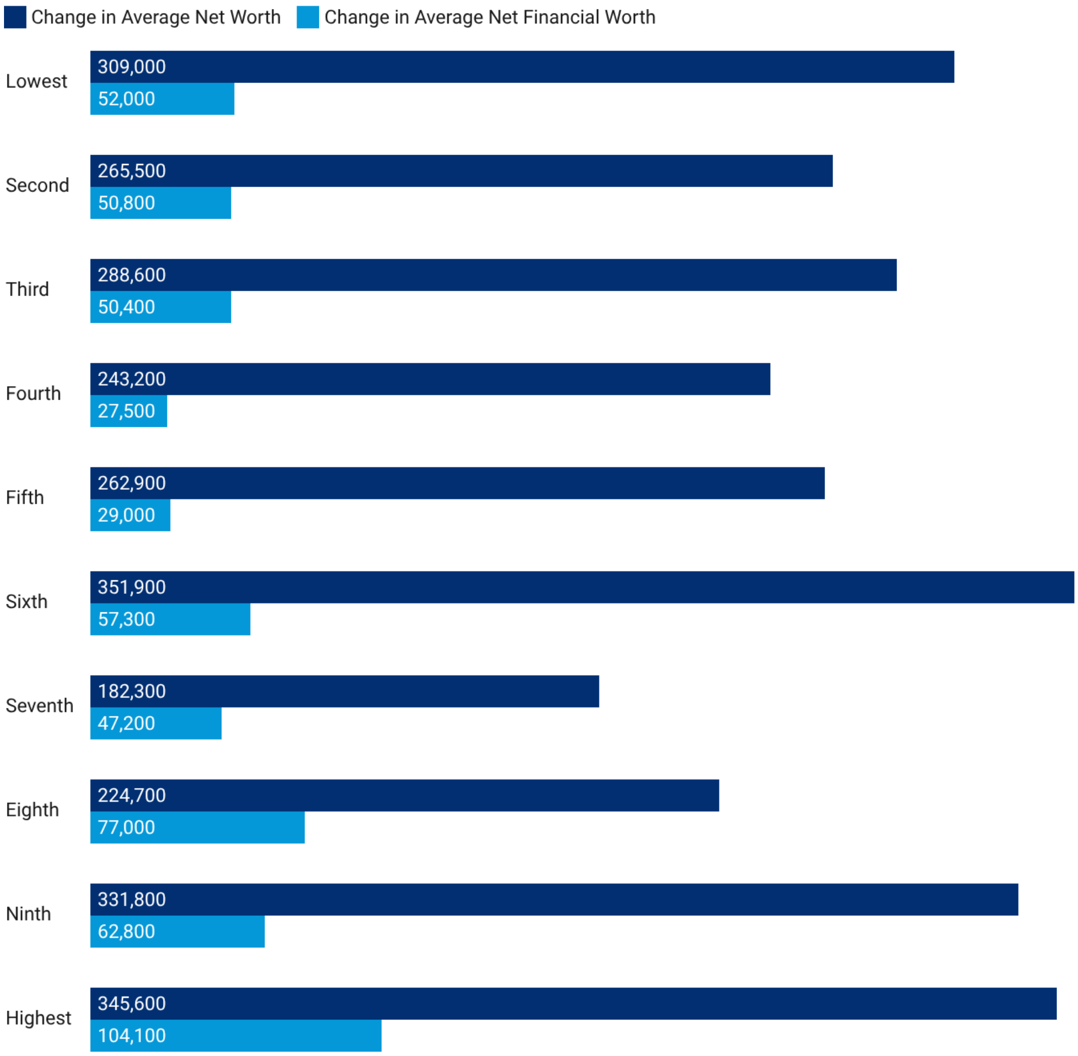
People have thus had harder times accumulating wealth for their retirement and for major purchases, such as housing or vehicles. More than two-thirds of Canadians are worried they will not have enough money for retirement according to various surveys conducted in 2025. Worrying that your future income will not be sufficient to cover your living costs fits the definition of cost of living concerns.

Moreover, economic research on subjective well-being, including work by Becker and Rayo (2007), demonstrates that what matters for happiness is how your real income compares to your expectations, and not your absolute real income level. Individuals and societies accustomed to faster real growth trajectories may experience deceleration as relative deprivation, even absent absolute decline. Current workers comparing their trajectory to their parents' experience, or to their own earlier career progression, may perceive themselves as falling behind despite objective improvement.

The proliferation of social media platforms has fundamentally altered financial

Figure 4: Change in Average Net Worth, by After-Tax Income Decile, 2012 - 2023

Constant dollar change over the period



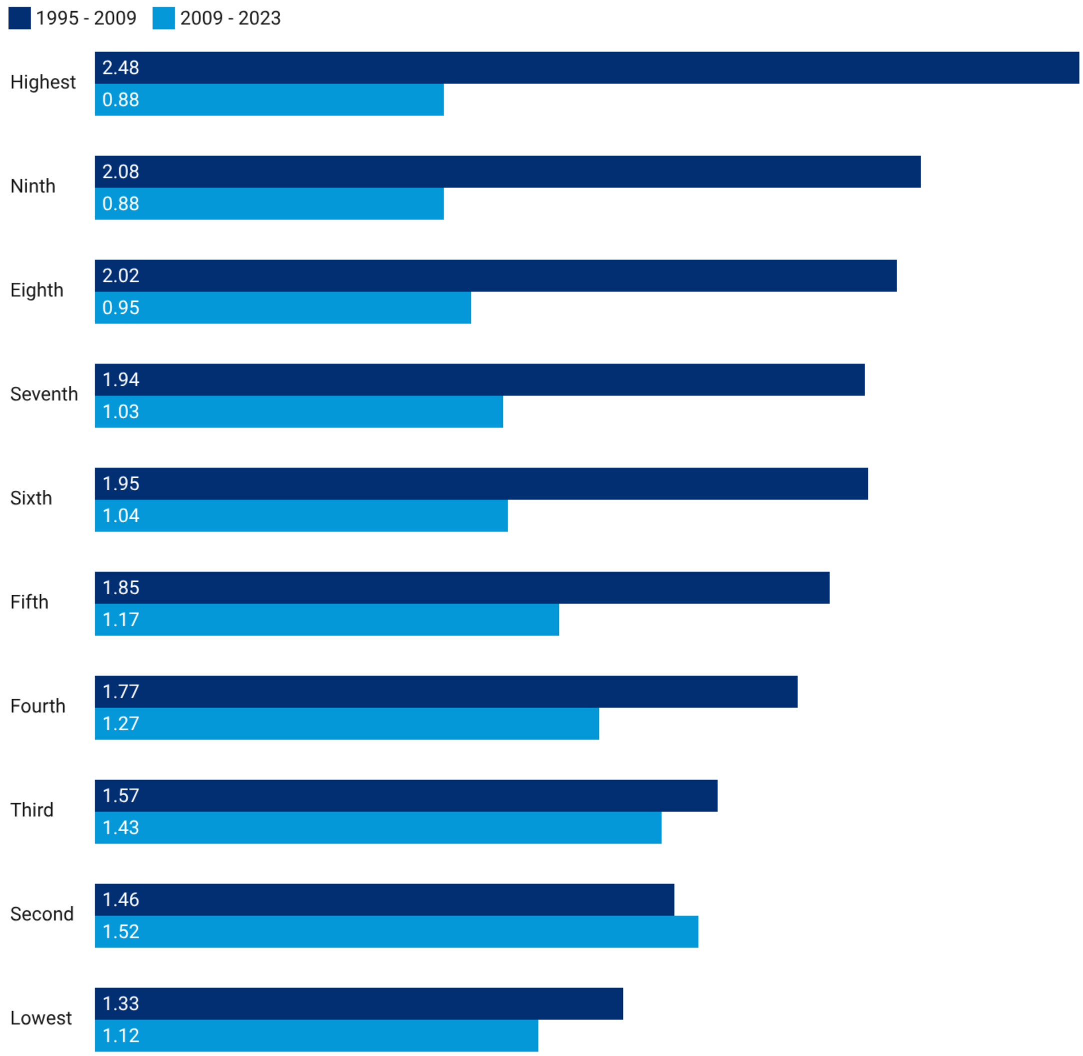
Source: Statistics Canada Table: 11-10-0078-01 • Created with Datawrapper

reference groups and consumption norms. Extensive research documents how social media creates unrealistic financial comparisons through curated content emphasizing luxury goods and aspirational lifestyles.

Humans are inherently social beings for whom relative social position matters significantly. Social media algorithms encourage users to perceive portrayed lifestyles as representative of peer groups, distorting perceptions of "normal" consumption and "necessary" expenditures. This can generate feelings of financial inadequacy, so-called financial dysmorphia, and pressure to emulate displayed lifestyles, potentially leading to overconsumption and financial anxiety. A study conducted by Intuit Credit Karma (2024) shows that more than 40 per cent of Gen Z and Millennial respondents experience money dysmorphia. A separate study by Hartford Funds found that 82 per cent of Americans who

Figure 5: Real Household Disposable Income per Decile

Average annual growth



Source: Statistics Canada: Table 11-10-0193-01 • Created with Datawrapper

experience money dysmorphia feel behind on their finances. The number of people using social media has quintupled since 2010, according to an analysis by Kepios. This suggests that increasing exposure to social media may contribute to perceived financial insecurity independent of actual economic circumstances.

3. Policy Evaluation and Recommendations

Political leaders across Canada and other developed nations have pledged to address cost of living concerns. Prime Minister Carney has made reducing costs and helping Canadians "get ahead" as a top priority. However, the analysis above suggests that different policy interventions vary dramatically in effectiveness, and some popular proposals may prove counterproductive.

3.1 Ineffective Approaches: Tax Cuts and Subsidies

Targeted measures such as Goods and Services Tax (GST) exemptions for first-time homebuyers may assist specific beneficiaries but generate negative externalities. By stimulating housing demand without addressing supply constraints, such policies increase prices for all market participants, potentially worsening overall affordability.

Broad-based tax reductions, such as general GST cuts, present similar challenges. First, only a portion of tax reductions translates into lower consumer prices; the remainder accrues as increased business profits, with the distribution determined by demand elasticity. Second, reduced government revenue necessitates either decreased public services and transfers — negatively affecting household real income — or increased public debt. Financing current consumption through higher public debt raises intergenerational equity concerns.

Government subsidies intended to create "well-paying jobs" and increase household income often rest on questionable economic logic. Subsidies to specific industries or regions are unlikely to create net employment; they merely reallocate workers from unsubsidized to subsidized sectors. This is because, in equilibrium, labour force expansion in preferred sectors requires drawing workers from elsewhere in the economy.

There is limited empirical evidence to suggest that subsidized firms treat employees better, share profits more equitably, or provide more skill development or higher-wage opportunities than other firms. Industrial policies increasing demand for higher-educated, better-compensated workers do not suddenly enable lower-educated workers to fill these positions or incentivize return to educational institutions.

3.2 Housing Affordability: Necessary but Difficult Reforms

Housing affordability challenges affect OECD countries broadly, with over 60 per cent of Canadians believing no level of government adequately addresses the issue. Deregulation, financialization, and government withdrawal from rental and affordable housing construction have increasingly transformed real estate from primarily shelter into primarily an investment vehicle.

Land fundamentally differs from other capital assets: it confers spatial monopoly power over an essential input for production and living. Unlike other capital, land is immobile, fixed in supply, and essential for diverse applications including housing, commerce, industry, and recreation.

House price appreciation in desirable urban areas results not from homeowner effort but from government investments and economic development making cities increasingly attractive. Despite not causing these value gains, homeowners capture financial benefits tax-free.

Economic theory suggests treating land analogously to natural monopolies or natural resource ownership: taxation, removal of entry barriers, and in some cases, potential nationalization. Countries implementing such approaches (such as Singapore to give one example) demonstrate superior housing affordability outcomes.

Widespread agreement exists that increased taxation on land and housing value appreciation would improve outcomes through two mechanisms: lowering prices by reducing speculative demand and increasing equity by redistributing gains to those without property wealth. However, political implementation faces significant obstacles when two-thirds of Canadians own homes purchased under assumptions of tax-free capital gains.

Direct government involvement in housing supply represents a critical intervention. Prior to 1980s reforms, governments in many countries invested heavily in public social housing. Canada's withdrawal from rental and affordable housing construction in the 1990s, intended to expand private market roles, produced drastic declines in investment for lower-income households. Over 80 per cent of existing social and affordable housing in Canada predates the mid-1990s.

Public housing offers distinct advantages: value appreciation accrues to taxpayers, and government can ensure reasonable rents and healthy tenant mixing. Research demonstrates that communities where low-income families interact with high-income families improve economic mobility for those born in poverty. The Build Canada Home initiative shows promise. The new federal agency aims to construct affordable housing at scale while supporting income diversity. Planned partnerships with municipalities, Indigenous communities, non-profit organizations, and co-operatives align with evidence-based practices. However, effectiveness and efficiency remain uncertain pending implementation.

Finally, governments should enhance competition among landowners through efficient transit system development and zoning law reform that eliminates incumbent protections.

3.3 Productivity Enhancement: The Fundamental Solution

Labour productivity improvement constitutes the primary long-run mechanism for increasing real income, and hence, enhancing perceived prosperity. More productive firms achieve lower production costs and prices, increasing product demand. Higher demand raises labour demand, boosting wages. Because of productivity gains, firms can afford higher wages without price increases, resulting in real wage growth.[1]

Productivity growth typically stems from organizational, procedural, or technical innovations rooted in research and learning. Innovation rates depend on numerous interacting factors including social context, cultural norms, educational systems, regulatory frameworks, and government support effectiveness. Few economists possess reliable prescriptions for accelerating long-term growth — it remains a "hard nut to crack."

However, comparative analysis can identify factors potentially explaining Canada's productivity underperformance relative to other countries, particularly the United States. Rosell *et al.* (2023) identify six key

1 For a more nuanced discussion of the long-term relationship between median wages and labour productivity in Canada, see Andrew Sharpe and James Ashwell (2021).

factors: small and dispersed markets, regulatory framework deficiencies, large presence of small and zombie firms, growing productivity gaps between frontier and non-frontier firms, skills and skill mismatches, and management education inadequacies. These factors can be grouped into two broad categories: weaker competitive environments and lower entrepreneurship skills.

International organizations and numerous Canadian commissions have proposed potential remedies, which can be categorized by implementation difficulty:

- Politically difficult but valuable policies include removing foreign investment limits (excluding national security concerns), shifting tax burden toward consumption, and eliminating ineffective business support programs. These face resistance because they harm specific constituencies while delivering benefits primarily in the long term.
- Design-challenging but worthwhile policies include making the Competition Bureau fully independent (Guttierrez and Philippon, 2018), reforming procurement systems (including eliminating industrial and technology benefits policies), addressing non-financial training barriers, improving merit-based educational support, enhancing business advisory services, facilitating access to government programs, and improving credential recognition.
- Empirically ineffective policies include subsidies, reducing environmental protections, and lowering corporate taxes. Canada ranks first among G7 nations on the International Tax Competitiveness Index, offers among the most generous research and development subsidies, and maintained lower effective marginal tax rates on investment than the United States for decades. Yet productivity continues lagging. Excessive subsidies for small- and medium-sized businesses enable poorly performing companies to survive, depressing overall productivity. When capital consumption allowances are generous, as in Canada, corporate taxes primarily capture economic rents without distorting investment decisions (Furno, 2021; Gechert and Heimberger, 2022). Research consistently finds that higher capital gains taxes do not harm economic performance and primarily affect high-income taxpayer savings.

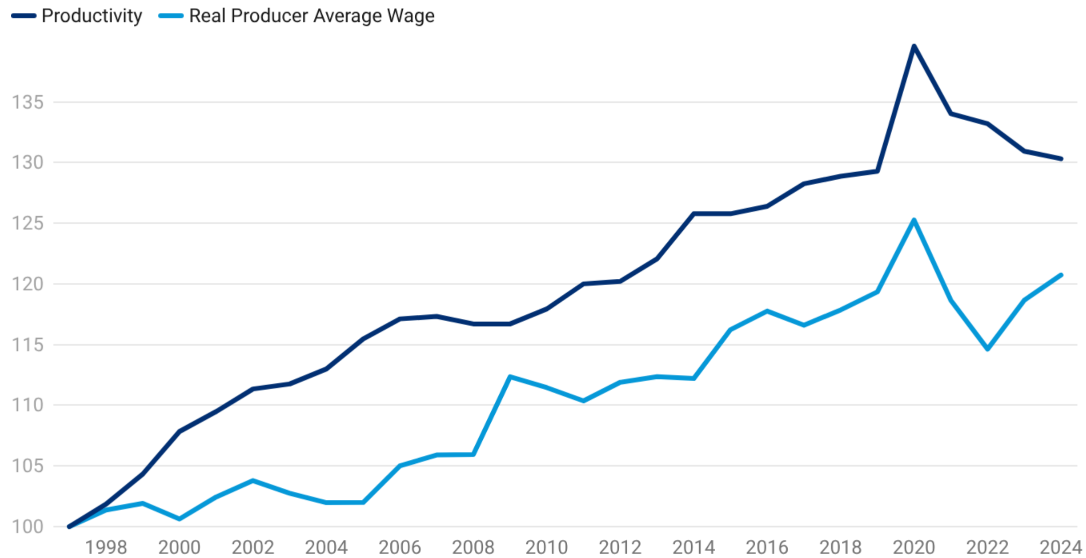
3.4 Ensuring Workers Benefit from Productivity Gains

In perfectly competitive markets, productivity increases translate directly into real wage growth. However, the Canadian economy has become increasingly concentrated, with certain firms controlling substantial market shares and employing large portions of workers in specific occupations or regions. These firms can earn significantly above-competitive profits and possess disproportionate bargaining power with employees. Productivity gains in such firms may therefore increase profits rather than wages or reduce prices.

Empirical evidence confirms this pattern (Figure 6). Between 1997 and 2007, labour productivity grew at 1.6 per cent annually while real average wages (deflated using producer prices) grew only 0.6 per cent annually. Real wage growth similarly lagged productivity growth following the Great Recession. Had real wages grown at productivity rates throughout these periods,

Figure 6: Labour Productivity and Real Wages

Indexed to 100 in 1997



Source: Statistics Canada Tables 36-10-0480-01, 14-10-0417-01, and 36-10-0130-01 • Created with Datawrapper

average real Canadian income would be 12 per cent higher—a meaningful improvement.

Ensuring Canadian households benefit from productivity gains requires strengthening worker bargaining power to prevent gains from accruing exclusively to domestic and foreign shareholders.

Economists, business leaders, and market-oriented enthusiasts since the Thatcher, Reagan, and Mulroney eras have accused unions of undermining economic growth and have systematically weakened workers' bargaining power. Private-sector unionization has plummeted from roughly 30 per cent in the early 1980s to just 15 per cent in 2023.

Interventions include facilitating unionization, streamlining certification processes, and enabling sectoral bargaining. While economic theory suggests unionization may harm growth, empirical evidence (Doucouliagos *et al.* 2017) reveals a different story: unions may negatively affect corporate profitability and investment but reduce earnings inequality, improve worker well-being, and demonstrate virtually no impact on productivity or economic growth.

Additional measures include ensuring all provinces adopt federal standards for mandated paid sick days, recognizing gig workers as employees, and implementing stronger anti-scab legislation.

3.5 Supporting Vulnerable Populations

Most Canadians expressing financial anxiety can nonetheless afford necessities and maintain decent living standards. However, some individuals face genuine hardship, confronting choices between feeding children and paying rent.

Creative destruction drives economic growth, and government should not impede this process. However, creative destruction disrupts existing structures, leaving some individuals negatively affected for extended periods. Consider workers in their mid-40s with twenty years of occupational experience who lose employment to artificial intelligence. Finding new employment at that age proves difficult, particularly when experience and skills relate to declining sectors. Research shows that laid-off workers who find new employment earn, on average, 10-20 per cent less than projected earnings in previous positions after five years.

Government support cushions economic disruption impacts and facilitates business and employee adaptation to shocks, limiting income variability. Canada's support generosity ranks low among OECD countries (OECD, 2018).

The current Employment Insurance (EI) system inadequately serves workers displaced by structural economic change. These workers typically maintained long employment histories before displacement and often earned above maximum insurable earnings. EI benefits may reduce income by more than half, while housing and food costs remain constant and employment prospects remain uncertain. Many countries have adopted benefit adjustments based on employment history, providing enhanced support for long-tenured workers.

Additional initiatives include personalized guidance and wraparound supports for long-tenured displaced workers, at-risk youth, single low-skilled adults, and other vulnerable populations.

4. Conclusion

High cost of living concerns dominate Canadian public opinion and appear consistently across many developed nations. This anxiety predates the COVID-19 pandemic, indicating structural rather than cyclical or country-specific causes.

Macroeconomic data reveal that, outside of the pandemic spike, inflation has generally remained near inflation targets, while most citizens experienced income growth exceeding cost of living increases. These aggregate statistics cannot explain widespread anxiety.

The evidence suggests three potential drivers: 1) declining housing affordability; 2) decelerating real income growth rates; and 3) social media-induced financial dysmorphia. While government cannot easily control social media usage, evidence-based solutions exist for housing affordability and income growth challenges.

Housing interventions should include increased government involvement in affordable housing supply, public transit investment, and zoning law reforms eliminating incumbent protections.Income growth interventions include removing foreign investment limits, shifting taxation toward consumption, eliminating ineffective business subsidies, reforming procurement systems, facilitating access to government programs, increasing Employment Insurance generosity for long-tenured workers, enhancing support for displaced workers and vulnerable populations, and facilitating unionization and sectoral bargaining.

These policies defy simple left-right political categorization. Economic prosperity requires open, competitive markets —

necessitating the removal of barriers, subsidies, and inefficient regulations. However, the economy also requires balanced bargaining power between workers and employers, support for those negatively affected by economic change, and government intervention where markets fail to deliver desired social outcomes.

Effective policy must address both the objective economic challenges facing specific populations and the broader subjective anxieties affecting most citizens, even those experiencing improved material circumstances. Only through this comprehensive approach can governments meaningfully respond to the cost of living crisis while promoting sustainable, equitable economic growth.

References

Becker, G. and L. Rayo (2007) "Habits, Peers, and Happiness: An Evolutionary Perspective," *American Economic Review*, pp. 487–491.

Credit Karma (2024) "Gen Z and Millennials Are Obsessed with the Idea of Being Rich, and It Could Be Leading to Money Dysmorphia," available at: https://www.creditkarma.com/about/commentary/gen-z-and-millennials-are-obsessed-with-the-idea-of-being-rich-and-it-could-be-leading-to-money-dysmorphia.

Hartford Funds (2025) "Money Dysmorphia: How Social Media Can Make Us Feel Behind Financially," available at: https://www.hartfordfunds.com/practice-management/client-conversations/financial-planning/money-dysmorphia-how-social-media-can-redefine-success.html.

Lagarde, C. (2014) "The Challenge Facing the Global Economy: New Momentum to Overcome a New Mediocre," Speech at Georgetown University, School of Foreign Service, October 2, International Monetary Fund, available at: https://www.imf.org/en/news/articles/2015/09/28/04/53/sp100214.

Doucouliagos, H., R. B. Freeman and P. Laroche (2017) *The Economics of Trade Unions: A Study of a Research Field and Its Findings*, Taylor and Francis.

Furno, F. (2021) "The Macroeconomic Effects of Corporate Tax Reforms," *arXiv Papers*, No. 2111.12799.

Gechert, S. and P. Heimberger (2022) "Do Corporate Tax Cuts Boost Economic Growth?" *European Economic Review*, Vol. 147.

Gutierrez, G. and T. Philippon (2018) "How EU Markets Became More Competitive than US Markets: A Study of Institutional Drift," *NBER Working Paper*, No. 24700.

Krugman, P. (2025) "Affordability, Part I," Substack, available at: https://paulkrugman.substack.com/p/affordability-part-i.

OECD (2018) *Income Redistribution Across OECD Countries: Main Findings and Policy Implications*, OECD Economic Policy Paper, No. 23.

Royal Bank of Canada (2025) "Financially Paralyzed: Higher Costs Have Canadians Feeling Unable to Move Forward – RBC Poll," January 23, available at: https://www.rbc.com/newsroom/news/article.html?article=125971.

Rosell, C., K. Dowsett and N. Paterson (2023) "A Critical Juncture: Assessing Canada's Productivity Performance and Future Prospects," *International Productivity Monitor*, No. 45.

Schembri, L. (2020) "Perceived Inflation and Reality: Understanding the Difference," Remarks to the Canadian Association of Business Economists, Bank of Canada, August 25.

Sharpe, A. and J. Ashwell (2021) "The Evolution of the Productivity-Median Wage Gap in Canada, 1976–2019," *International Productivity Monitor*, No. 41 (Fall).