

Ceteris Paribus

Is it time for a wealth tax?

Regulating prediction markets

The great productivity
debate

OCTOBER 2026



CONTENTS

In this edition

From the Editors-in-Chief	3
Daniel Whitehouse and Henrik Helsen	
What Peacocks Know That Career Advice Doesn't	4
Riddhi Bhansali	
"Manchesterism": Will Andy Burnham Bring an End to UK Regional Inequality?	7
Simon McKim	
Would a Wealth Tax Work?	13
Toby Page	
War, Oil and the Environmental Price of Geopolitics	18
Aditya Biju	
Contracts No Regulator Has Graded: The Information Question America Answered Blind and Britain Never Asked	25
Dan Low	
The US-EU Productivity Paradox: The Debate Around Measurements	37
Sohan Sahu	

FRONT MATTER

From the Editors-in-Chief

What you see on your screen has been a long time in the making. Over the past nine months, working in collaboration with Warwick Economics Society, we are proud to bring the fine-tuned analysis of some of this University's best writers to a broader audience. In these pages, you'll find some of the thorniest economic problems of this extraordinarily turbulent time dissected and simplified in a way that is accessible to all those with an interest in economics. Yet we move beyond the superficial soundbites we often see in the media to engage with tough problems using real data and statistical analysis.

We hope that this publication offers something a little different from the usual extracurricular Ferris wheel of ephemeral contributions. Economics and its related disciplines touch almost every aspect of our lives. When competing opinions surface, as they so often do, we believe there is value in doing what a university education ought to equip us for: interrogating complex ideas by rigorously applying the analytical tools at our disposal.

We are grateful to EconSoc for supporting this project every step of the way. Our particular thanks go to Ryan Gonsai and Kelly Ng for their invaluable help in bringing this first edition to life. The fun doesn't stop here, however – in future editions, you (yes, that's you, reading this) will have the opportunity to be considered for inclusion with articles of your own choice. Keep your eyes peeled...

Happy reading,

Daniel Whitehouse

Henrik Helsen

Our team

Editors-in-Chief

Daniel Whitehouse
Henrik Helsen

Contributing Writers

Riddhi Bhansali
Aditya Biju
Simon McKim
Toby Page
Sohan Sahu

Head of Research and Data

Dan Low

What Peacocks Know That Career Advice Doesn't

Riddhi Bhansali

A peacock's tail is, from an engineering standpoint, a bad idea. It's heavy, it's bright, it takes real metabolic effort to grow, and it makes the bird slower and easier for a predator to catch. Any sane designer would cut it. Evolution didn't.

The biologist Amotz Zahavi worked out why in the 1970s, in what's now called the handicap principle (Zahavi 1975). The tail is credible because it's costly. A weak, sickly bird couldn't afford to grow one and still survive. So the extravagance itself is the proof of fitness: anyone can fake a cheap signal, but you can't fake an expensive one.

The same logic runs through things we build, not just things we evolve. Would a university degree hold much credibility if it didn't cost a lot of money, time, or effort? Would a city like Dubai attract as many tourists as it does if it didn't have the Burj Khalifa? No. In both cases, the cost isn't overhead. It's the point. To be respected, you need to put in real costs: money, effort, time. The expense isn't what you pay despite the signal being valuable. It's what makes the signal valuable in the first place.

This is one of the more powerful things to have migrated from biology into economics, and it explains something that has always bothered me about the culture of self-improvement: the popular injunction to "fake it till you make it" is just a request to send a costless signal in a market that only rewards costly ones. It cannot work for long, and understanding why requires taking the handicap principle seriously as economic theory rather than a TED speech.

In 1973, the economist Michael Spence published a paper on the job market and arrived at the same conclusion Zahavi had reached independently in biology (Spence 1973). Spence's puzzle was this: employers cannot directly observe a candidate's productivity before hiring them, so they look for a proxy. Education turns out to be a remarkably good one, not because calculus is used on the job, but because acquiring a degree is expensive. A highly able candidate finds the years of study, the exams, and the discipline required comparatively easy to bear while a less able candidate finds the same ordeal punishing, expensive, and more likely to end in failure, which results in a fine line differentiating them.

What makes this Spencerian result identical in structure to Zahavi's biological one is the requirement of a single crucial condition: the marginal cost of producing the signal must fall as underlying quality rises. It is the reason both the peacock's tail and the diploma

act as honest signals, and it is the reason a signal that fails this condition, one that is ‘low effort’ for everyone to produce, tells the observer nothing at all.

Confidence fails this condition completely. Anyone, competent or not, can adopt an assured posture, speak fluently in a pitch meeting, or claim mastery on a resume. The cost of appearing confident does not scale inversely with actual ability the way the cost of a peacock’s tail scales with actual vigor. It is, in the vocabulary of game theory, cheap talk. It is costless and has non-binding communication that carries no guarantee of truthfulness.

The formal treatment of this problem belongs to Vincent Crawford and Joel Sobel, whose 1982 model of strategic information asked when cheap talk can nonetheless be informative (Crawford and Sobel 1982). Their answer was narrow. Cheap talk only conveys useful information when the interests of the sender and receiver are sufficiently aligned. Where they diverge, as between a founder and an investor, or an employee and a manager, the incentive to exaggerate overwhelms the incentive to inform, and the equilibrium collapses into what they call babbling, talk so detached from truth that a rational listener optimally ignores it.

Akerlof’s 1970 lemons problem describes the same collapse from the other side (Akerlof 1970). Sellers of lemons know more about quality than buyers do, and lacking any costly way to prove it, buyers rationally assume the worst. Good cars cannot cheaply distinguish themselves from lemons and are driven from the market, not because they stop existing, but because there is no credible signal left to prove they do. The pattern is identical wherever unverified conviction floods a market, a pitch deck, a resume, an admissions essay. Once faking conviction becomes costless and universal, conviction stops carrying power, and observers correctly discount it towards zero.

The interesting economics begins once we ask what a credible signal looks like instead. Alan Grafen’s 1990 formalization of the handicap principle, which put Zahavi’s original argument on rigorous mathematical footing, showed that a signal is evolutionarily stable only if it is costly enough that mimicking it would be irrational for a lower-quality sender (Grafen 1990). This is the case for high-cost advertising: only firms with quality products can afford it since they are the ones earning a higher revenue from repeat customers to recoup this costly decision (Nelson 1974; Milgrom and Roberts 1986).

The same logic explains why dividends move stock prices more than a glowing press release does (Bhattacharya 1979). A firm’s declaration that business is strong is baseless. A dividend paid out is real talk since only a firm confident in sustained future earnings can afford to make that payment repeatedly without straining itself.

None of this is an argument against ambition or self-belief as private psychological states. It is an argument against treating the outward performance of belief as a substitute for the costly evidence that markets and employers have been designed to demand. Biology enforces honesty through cost. Fake it until you make it asks an audience to accept a false signal. Sometimes this works. In markets where nobody is looking closely, or where events move fast enough that no one has time to check, a bluff can carry the day, the same way a weak poker hand occasionally wins because the other player folded first. A

fabricated internship on a resume can survive a first screening. It can even survive an interview, if the interviewer asks only the questions a confident answer can deflect.

But the handicap principle does not stop at the interview. It governs what happens afterward, when the job actually starts. A candidate who never did the work they claimed has no real model of how the department functions, no memory of the specific problems it faces, no instinct for its rhythms, because none of that can be attained from confidence itself. Thus, when the gap between what they claim and what they know becomes visible, which it always does under the sustained scrutiny of daily work, they are let go. What remains standing afterward is never the performance. It is only ever the cost that stood behind it.

References

- Akerlof, G. A. (1970). “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism”. In: *The Quarterly Journal of Economics* 84.3, pp. 488–500. DOI: 10.2307/1879431.
- Bhattacharya, S. (1979). “Imperfect Information, Dividend Policy, and ‘The Bird in the Hand’ Fallacy”. In: *The Bell Journal of Economics* 10.1, pp. 259–270. DOI: 10.2307/3003330.
- Crawford, V. P. and J. Sobel (1982). “Strategic Information Transmission”. In: *Econometrica* 50.6, pp. 1431–1451. DOI: 10.2307/1913390.
- Grafen, A. (1990). “Biological Signals as Handicaps”. In: *Journal of Theoretical Biology* 144.4, pp. 517–546. DOI: 10.1016/S0022-5193(05)80088-8.
- Milgrom, P. and J. Roberts (1986). “Price and Advertising Signals of Product Quality”. In: *Journal of Political Economy* 94.4, pp. 796–821. DOI: 10.1086/261408.
- Nelson, P. (1974). “Advertising as Information”. In: *Journal of Political Economy* 82.4, pp. 729–754. DOI: 10.1086/260231.
- Spence, M. (1973). “Job Market Signaling”. In: *The Quarterly Journal of Economics* 87.3, pp. 355–374. DOI: 10.2307/1882010.
- Zahavi, A. (1975). “Mate Selection—A Selection for a Handicap”. In: *Journal of Theoretical Biology* 53.1, pp. 205–214. DOI: 10.1016/0022-5193(75)90111-3.

“Manchesterism”: Will Andy Burnham Bring an End to UK Regional Inequality?

Simon McKim

Since the decline of traditional northern industries like coal, steel and shipbuilding in the 1980s, along with a boom in southern professional services, concerns have been raised about Britain as a two-sided coin – a booming south and a crumbling north. In recent years, politicians have campaigned on regional UK inequalities, examples being Boris Johnson’s ‘Levelling Up’ campaign (HM Government 2022), Keir Starmer promising to end educational inequality (The Labour Party 2023) and most recently Andy Burnham, the former mayor of Greater Manchester and now Prime Minister, who has pledged to bring the successes of ‘Manchesterism’ and regional devolution to the whole country (Prime Minister’s Office 2026). These policies are mentioned with good reason. Regional inequalities can foster resentment and apathy towards politics, whilst leaving communities unproductive and in poverty. Undoubtedly, inequalities still exist. The GDP per capita of London remains 75% larger than the UK average and 140% larger than the UK’s poorest region (Office for National Statistics 2025). The percentage of children on free school meals is 12 percentage points higher in the North East than the South East (Department for Education 2025). Many of the highest-paying jobs still cluster around London, creating a brain drain of resources away from poor areas. However, the dynamics of regional disparities are changing. With the impressive growth of certain cities outside London, such as the economic booms of Manchester, Liverpool and Leeds, a period of monopoly over economic power may be ending. I plan to look at whether other regions are truly beginning to catch up, or at least to stop falling further behind.

Firstly, I analysed regional GDP growth figures since 1999 using ONS local authority data (Office for National Statistics 2025). To test if growth rates have converged between regions, I divided the data into two equal-length periods, 1999–2011 and 2012–2023. The results follow:

Results show clear changes amongst several regions. UK growth from 1999–2011 was dominated by London, Scotland and the South East, with London a huge outlier as the post-deregulation financial services industry rose. Since 2012, however, the North West and the East of England have risen to become the fastest-growing regions, ahead of London. Splitting the English regions into groups and taking simple averages of the figures in the table shows that the gap between them has narrowed. The South’s average growth rate (South East, London, South West and East of England) fell from 0.0196 to 0.0169, while the North’s (North West, North East and Yorkshire) was essentially flat, at 0.0156 and then 0.0155, and the Midlands’ rose slightly, from 0.0132 to 0.0136. The convergence therefore comes mainly from the South slowing down rather than from the North catching

Region	1999–2011 Average Growth	2012–2023 Average Growth
South East	0.0184	0.0147
London	0.0276	0.0179
South West	0.0159	0.0163
East of England	0.0163	0.0186
East Midlands	0.0149	0.0128
West Midlands	0.0115	0.0144
Yorkshire and the Humber	0.0143	0.0160
North West	0.0160	0.0190
North East	0.0166	0.0115
Wales	0.0153	0.0127
Scotland	0.0197	0.0118
Northern Ireland	0.0134	0.0144

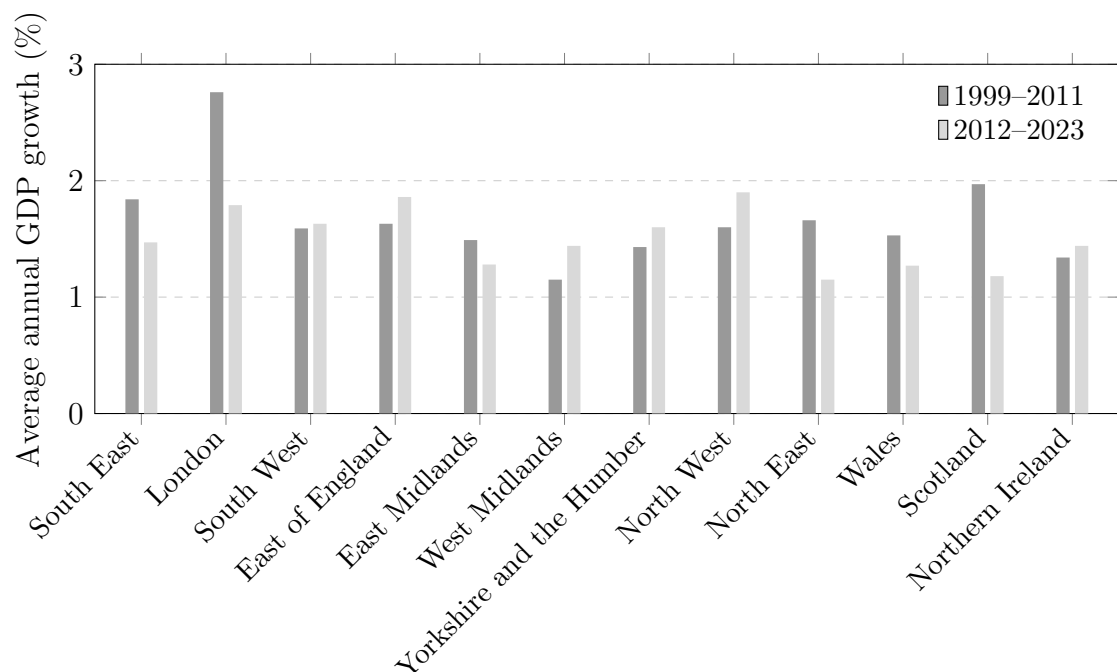


Figure 1: Average annual GDP growth by UK region, 1999–2011 versus 2012–2023 (figures from the table above).

up. The North West, West Midlands and East of England showed the largest increases in growth, with Yorkshire also improving. These data points provisionally suggest that UK economic power may be becoming less concentrated in London and the South East, even if the North as a whole has not yet caught up. That being said, the UK's poorest regions, Wales and the North East, still have substantially lower growth rates than the UK average.

Which industries is growth coming from?

To investigate these results, I looked at UK data by industry. In the regions that have seen increases in growth – the North West, West Midlands and Yorkshire – traditional industries of coal, steel and heavy manufacturing declined across the late 20th century. I have examined whether, as these regions have seen economic gains, these gains have come from increases in high-skilled industries. Much of the economic expansion seen in London has occurred due to growth of financial and legal services and other high-skilled jobs. By measuring industry, I can see if the same approach is being attempted in these regions. With ONS industry data (Office for National Statistics 2025), I ranked the industries in each of these regions by their contribution to economic growth over the past ten years. The results follow for the five most contributory industries by region:

North West

	Percentage of total economic growth
Professional, scientific and technical activities	13.8%
Information and communication	12.6%
Real estate activities	9.5%
Manufacture of transport equipment	8.1%
Construction	7.9%

Yorkshire and The Humber

	Percentage of total economic growth
Information and communication	14.4%
Professional, scientific and technical activities	10.2%
Construction	10.1%
Real estate activities	9.7%
Wholesale and retail trade; repair of machinery	5.8%

West Midlands

	Percentage of total economic growth
Information and communication	15.8%
Real estate activities	11.5%
Construction	9.6%
Professional, scientific and technical activities	9.4%
Education	7.0%

These results appear to reveal that a broad range of industries is fuelling the growth increases across these regions. Large contributions to growth were seen both in professional services, such as information and communication, and in sectors such as construction and the manufacture of transport equipment. This suggests growth may be more diversified and less reliant on a single industry. This diversity contrasts with the London economic boom of the late 20th and early 21st centuries, which was driven to a far greater extent by financial and insurance services.

Overall, the data I have collected suggests the UK may be becoming less reliant on London for economic growth. Other regions have shown increases in their rates of growth, whilst maintaining a diversified economy. Whether this is repeatable across the whole UK is a difficult question. I will look at Manchester as an example of economic success outside London, for two main reasons. Firstly, Greater Manchester has been one of the fastest-growing UK city regions, averaging 3.1% annual growth since 2015 (Greater Manchester Combined Authority 2025). Secondly, the UK Prime Minister, as former mayor of Greater Manchester, has talked extensively about emulating his successes there across the country.

The Growth of Manchester

Like much of the North West, Manchester's GDP growth over the last 10 years has been fuelled by professional and scientific services, as well as construction, communication and education (Office for National Statistics 2025). Many areas of Greater Manchester have also seen dramatic improvement through regeneration projects. The Bee Network, first set out under Burnham in 2018 (Greater Manchester Combined Authority 2020), has created a unified bus, tram and cycling network. Public transport has been expanded with developments like the Trafford Park tram line (Railway Gazette International 2020). City centre regeneration projects at Mayfield, First Street and Great Jackson Street have all improved and revitalised central Manchester (Manchester City Council n.d.).

Manchester has also cultivated a young and highly educated population: in the mid-2010s, 51% of the city's university students stayed in Manchester for work after graduating (Centre for Cities 2019). A young workforce in the city has attracted employers as well, with firms such as BNY and Vanguard expanding into Manchester in recent years (Haynes 2023; Invest Manchester 2024). Another mayoral initiative is the Good Growth Fund, a £1 billion investment scheme leading dozens of projects improving housing and employment across Greater Manchester (Greater Manchester Combined Authority 2025).

All of this leads to the obvious question of how the so-called 'Manchesterism' policy could be applied at either a regional or national level. An obvious step is further devolution to

regional authorities. Burnham has announced plans for regional mayors to keep a share of income tax and business rates, allowing further spending by localities on public services (Prime Minister’s Office 2026). Expansions of local control could increase funding for other transport networks and inner-city projects. Burnham has also focused on public transport, returning to the £2 bus fare cap he maintained in Greater Manchester (ITV News 2026). However, the UK is a diverse country. Attempts to apply what has worked in Manchester to all regions may be more difficult than imagined. Critics have already warned that giving mayors a share of income tax will benefit wealthy regions at the expense of poorer ones (Sandford 2026). Burnham’s ‘Manchesterism’ is likely to face hurdles.

References

- Centre for Cities (2019). *The Great British Brain Drain: An Analysis of Migration to and from Manchester*. R. McDonald. URL: <https://www.centreforcities.org/reader/the-great-british-brain-drain-an-analysis-of-migration-to-and-from-manchester/a-more-detailed-look-at-the-movement-of-students-and-graduates/> (visited on 26 September 2026).
- Department for Education (2025). *Schools, Pupils and Their Characteristics: Academic Year 2024/25*. Explore Education Statistics. URL: <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/2024-25> (visited on 26 September 2026).
- Greater Manchester Combined Authority (2020). *Change a Region to Change a Nation: Greater Manchester’s Walking and Cycling Investment Plan*. URL: https://assets.ctfassets.net/nv7y93idf4jq/Xx5s7azQY1SYmdNKIAviX/e4395ab029410907365cd0962d17bf81/19-1950_Bee_Network_delivery_plan-style_-_website_version.pdf (visited on 26 September 2026).
- (2025). *A Decade of Good Growth in Greater Manchester*. URL: <https://www.greatermanchester-ca.gov.uk/news/a-decade-of-good-growth-in-greater-manchester/> (visited on 26 September 2026).
- Haynes, L. (2023). *Manchester Offices Scramble: The Firms Lining Up 100,000-Square-Foot-Plus Moves*. CoStar. URL: <https://www.costar.com/article/1050347271/manchester-offices-scramble-the-firms-lining-up-100000-square-foot-plus-moves> (visited on 26 September 2026).
- HM Government (2022). *Levelling Up the United Kingdom*. CP 604. URL: https://assets.publishing.service.gov.uk/media/61fd3ca28fa8f5388e9781c6/Levelling_up_the_UK_white_paper.pdf (visited on 26 September 2026).
- Invest Manchester (2024). *Case Study – Vanguard*. URL: <https://www.investmanchester.co.uk/success-stories/case-study-vanguard/> (visited on 26 September 2026).
- ITV News (2026). *Andy Burnham Says £2 Bus Fare Cap ‘Will Put Money in People’s Pockets’*. URL: <https://www.itv.com/news/2026-07-22/andy-burnham-unveils-2-bus-fare-cap-in-pledge-to-be-cost-of-living-government> (visited on 26 September 2026).

- Manchester City Council (n.d.). *City Centre Growth and Infrastructure*. URL: <https://www.manchester.gov.uk/planning-and-regeneration/regeneration/city-centre-growth-and-infrastructure/city-centre-regeneration-areas> (visited on 26 September 2026).
- Office for National Statistics (2025). *Regional Economic Activity by Gross Domestic Product, UK: 1998 to 2023*. URL: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/bulletins/regionaleconomicactivitybygrossdomesticproductuk/1998to2023> (visited on 26 September 2026).
- Prime Minister's Office (2026). *PM Hands Mayors Share of Income Tax to Make Lives Better in Every Postcode*. GOV.UK. URL: <https://www.gov.uk/government/news/pm-hands-mayors-share-of-income-tax-to-make-lives-better-in-every-postcode> (visited on 26 September 2026).
- Railway Gazette International (2020). *Manchester Metrolink Opens Trafford Park Line*. URL: <https://www.railwaygazette.com/projects-and-planning/manchester-metrolink-opens-trafford-park-line/56080.article> (visited on 26 September 2026).
- Sandford, M. (2026). *Could Andy Burnham Let Mayors Raise More Taxes?* House of Commons Library. URL: <https://commonslibrary.parliament.uk/research-briefings/cbp-10937/> (visited on 26 September 2026).
- The Labour Party (2023). *Keir Starmer Unveils Labour's Mission to Break Down Barriers to Opportunity at Every Stage*. URL: <https://labour.org.uk/updates/press-releases/keir-starmer-unveils-labours-mission-to-break-down-barriers-to-opportunity-at-every-stage/> (visited on 26 September 2026).

Would a Wealth Tax Work?

Toby Page

Introduction

Wealth taxes have been a contentious topic of discussion over the past few years in academic and political debates, promoted by supporters as a tool to raise greater revenue for the government and reduce wealth inequality. There are plenty of philosophical and practical arguments for and against wealth taxes, with Adam and Miller (2021) providing a useful summary. This article is not an ethical discussion, but rather an analysis of the practical difficulties involved in implementing a wealth tax. Issues of capital flight, high administrative costs, and the difficulty of valuation are all major implementation challenges. Using historical examples as well as current evidence, this article demonstrates that these implementation problems are mostly solvable with smart policy design.

Whether a wealth tax should be implemented is beyond the scope of this article; the narrower question is whether it could work in practice. Wealth taxes could work. However, the political realities prevent an effective wealth tax from being implemented in the United Kingdom in the current climate. A successful wealth tax must prevent capital flight, which the UK is unable to achieve.

Before jumping into the debate, the discussion is based on an annual wealth tax on assets above a certain threshold: most proposals are 1% or 2% on assets over £10 million. These types of wealth taxes enjoy strong public support (Tax Justice UK 2025). It is also worth noting that the UK already taxes wealth in several ways, including through inheritance tax, capital gains tax, council tax and stamp duty (Rutter and Tetlow 2025). Where the term ‘wealth tax’ is used below, it refers to an annual tax on net assets above a high threshold.

Administrative Costs and Valuation Difficulties

One common objection is the administrative cost and complexity of implementing a wealth tax. The argument is that many assets are difficult to value objectively, such as private companies or art, and the administrative cost of valuing individuals’ wealth every year would severely limit the utility of a wealth tax. In fact, valuation issues were the main reason that the Wealth Tax Commission recommended against an annual, recurring wealth tax (Advani et al. 2020). A wealth tax may also cause a shift towards hard-to-value assets in attempts to avoid the tax, creating distortions in the economy. For example, private companies may resist IPOs to prevent the government from easily attaining a value of the owners’ assets.

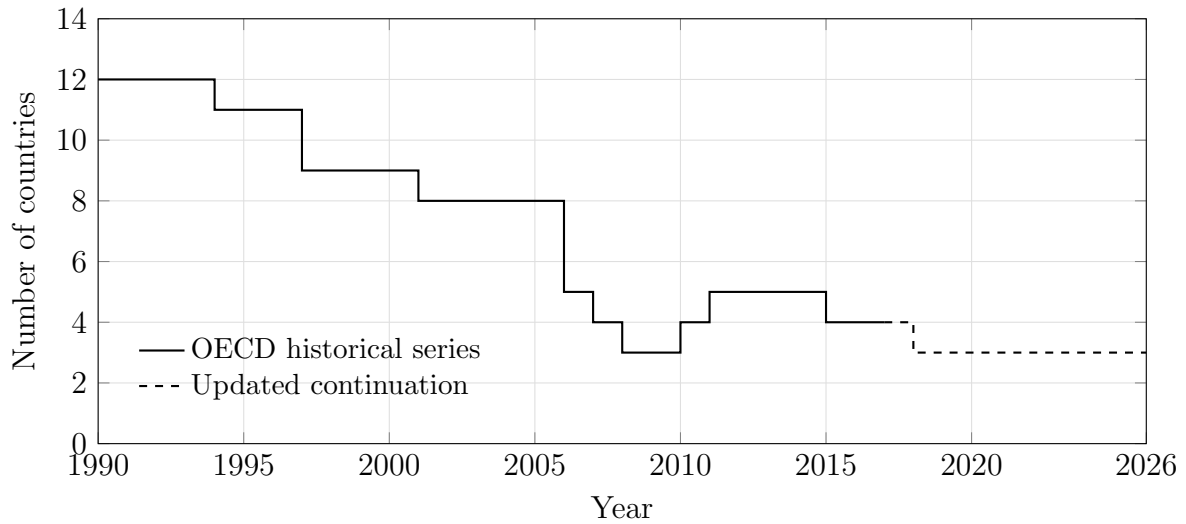


Figure 1: Number of European countries levying broad-based individual net wealth taxes, 1990–2026.

Source: (OECD 2018a, 2018b; Enache 2026)

There are solutions to these problems. Firstly, the difficulty of valuation is refutable. As proposed wealth taxes often cite £10 million as a minimum threshold, taxable assets would be large in value. Highly valuable assets are very likely to already have valuations. For example, expensive art would be insured by any sensible collector, with an insurance value attached to each painting. Private companies can also be valued using formulas, as done for Switzerland’s wealth tax, if they have not already been valued. Furthermore, a high threshold like £10 million means relatively few individuals will be subject to the tax, reducing the administrative costs of the wealth tax.

This point links closely to the design of the tax base. A large number of exemptions raises the complexity and thus administrative and compliance costs of any wealth tax. This situation was the case with historical wealth taxes, where exemptions were rife. A broad base with few exemptions would reduce the administrative burden of implementing the tax. Administrative and valuation difficulties are not strong enough to negate the efficacy of a wealth tax.

Historical Examples – Why Did Past Wealth Taxes Fail?

Twelve European OECD countries levied individual net wealth taxes in 1990, with most being repealed during the 1990s and 2000s. As of 2026, just three European countries continue to levy broad-based individual net wealth taxes: Spain, Norway, and Switzerland. Opponents of wealth taxes have claimed that this evidence proves the futility of attempting to implement a wealth tax. However, previous attempts failed largely due to design issues and the raft of exemptions, which enabled easy avoidance. Many exemptions were ostensibly well-intentioned; for example, business assets were exempt to limit burdens on productive activity. Moreover, tax caps aimed to limit liquidity issues, where the amount owed was limited to a proportion of an individual’s taxable income.

The issue with exemptions is the erosion of the tax base. Revenue generated was reduced,

higher complexity was created leading to higher administrative costs, and the wealth taxes became very easy to avoid for the super-rich. In fact, many exemptions were enacted through lobbying by the wealthy, who used them to quietly avoid the tax. As a result, the taxes became less effective, strengthening calls for their repeal. Efforts by the super-rich to change the framing of wealth taxes have also been relatively successful in reducing support for wealth taxes on philosophical grounds. Thus, abandonment of wealth taxes in the past does not reflect a failure of the principle of wealth taxes, but rather the success of special interests.

Perret (2021) concludes that the political economy channel was the main reason for the repeal of most wealth taxes, as it also influenced the number of exemptions which heavily reduced the efficacy of wealth taxes. Historical wealth taxes raised far less money than a simple 1% tax on the top 1%, despite broader coverage, largely because of exemptions, design issues and the ensuing ease of avoidance. Therefore, the failure of previous wealth taxes is not a valid argument against wealth taxes today. The case for wealth taxes is actually even stronger today, with greater ability to tax wealth coming from greater international tax transparency and digitisation. The perceived failure of historical wealth taxes demonstrates that designing wealth taxes is key to their success.

Proponents further argue that previous wealth taxes exhibited little behavioural change. Exaggerations of an exodus of the wealthy and a collapse of investment did not occur. Nonetheless, the lack of behavioural change likely stemmed from the ease of avoidance. There would likely be greater behavioural change if avoidance channels were closed down. A properly designed wealth tax would likely cause noticeable changes (Advani and Tarrant 2021), such as capital flight, where owners of capital exit a country in response to policy changes.

Capital Flight

Capital flight is perhaps the most common objection raised with regard to wealth taxes, especially when proposed in only one country. For example, a small increase in the Norwegian wealth tax in 2023 coincided with a record number of multimillionaires leaving the country (Neate 2023). Many also argue that the super-wealthy are incredibly mobile and that changing citizenship would cause very little inconvenience to their lives.

However, proponents of wealth taxes have hit back at such objections, chiefly citing that very few wealthy individuals migrate for tax reasons. Migration of the wealthy is driven by many factors rather than specific tax policies, and talk of capital flight is often a PR stunt to prevent the imposition of higher taxes on the wealthy. Some wealthy individuals state that they would not move to tax-free districts such as the Bahamas because it would be boring to live there for more than a few weeks (Friedman et al. 2024).

Many scholars do, however, understand that the international super-wealthy are very mobile and can easily change citizenship to avoid tax. They contrast with homegrown multimillionaires, who are much less likely to migrate. Whilst capital flight is usually small relative to the increase in revenue as a result of a wealth tax (Tax Justice UK 2025), it does pose a problem (Advani et al. 2024). The loss of capital reduces a country's investment levels, depressing economic growth and likely causing a decline in employment

and national income.

French economists Emmanuel Saez and Gabriel Zucman, well-known proponents of taxation on wealth, argue that exit taxes and ‘tail’ provisions are two ways of avoiding the capital flight issues associated with tax competition (Saez and Zucman 2022). ‘Tail’ provisions are where people are taxed for 5 or 10 years after moving to another country such that they cannot avoid tax. Most G7 countries do in fact have exit taxes in some form, and this policy would certainly reduce the capital flight issue. The United States taxes citizens wherever they are located, the only major economy to practise such citizenship taxation.

The problem with these solutions is the political difficulty. Any whiff of an exit tax may prompt migration by the wealthy to pre-empt any such action, and the UK would find it difficult to impose citizenship taxation as it would require the renegotiation of its over 100 international tax treaties. It is possible to solve the capital flight issue with policy design, but enacting the necessary policies would be too politically challenging in the UK’s case. A final thought is that those who would leave are likely to have already left, as British taxes are already much higher than in tax havens. However, this conjecture is certainly not enough to base tax policy on. A wealth tax would raise revenue, but without exit taxes or citizenship taxation the UK would have little defence against the departure of its highest-value taxpayers, who account for most of the wealth that leaves (Advani et al. 2024). Even a small number of departures could therefore outweigh the revenue gained and cause greater harm to the UK economy.

Crowding Out Tax Reform

Of course, calls for a wealth tax can overshadow reform of existing taxes. Many commentators prefer these reforms before a wealth tax is even thought about, as they are a simpler way of generating revenue (Leunig 2026). Clearly, any reforms to tax that make the system more efficient are beneficial, and reforming capital gains or inheritance tax appears to be a simpler route to improving the taxation of wealth. It is understandable that talk of a wealth tax may crowd out desire for reform of other existing taxes on wealth. The reform of existing taxes should certainly be carried out before a wealth tax, as reform is less costly and an ‘easier win’.

Conclusion

Overall, the main arguments against wealth taxes have been shown to be solvable, and the failure of historical wealth taxes in Europe reflects a success of special interests instead of a failure of the idea. Despite these solutions, implementation in the UK is unfeasible, largely due to the UK’s inability to prevent capital flight. Furthermore, reform to existing taxes is far easier to achieve and can achieve many of the same goals.

References

Adam, S. and H. Miller (2021). “The Economic Arguments For and Against a Wealth Tax”. In: *Fiscal Studies* 42.3–4, pp. 457–483.

- Advani, A., E. Chamberlain, and A. Summers (2020). *A Wealth Tax for the UK: Final Report*. Wealth Tax Commission. URL: <https://www.lse.ac.uk/International-Inequalities/Assets/Documents/WealthTaxCommission-Final-report.pdf> (visited on 1 July 2026).
- Advani, A., A. Summers, and C. Poux (2024). *Business Owners Who Emigrate: Evidence from Companies House Records*. Centre for the Analysis of Taxation Policy Brief. URL: https://centax.org.uk/wp-content/uploads/2024/10/AdvaniPouxSummers2024_BusinessOwnersWhoEmigrate.pdf (visited on 1 July 2026).
- Advani, A. and H. Tarrant (2021). “Behavioural Responses to a Wealth Tax”. In: *Fiscal Studies* 42.3–4, pp. 509–537.
- Enache, C. (2026). *Wealth Taxes in Europe, 2026*. Tax Foundation Europe. URL: <https://taxfoundation.org/data/all/eu/wealth-taxes-europe/> (visited on 13 September 2026).
- Friedman, S. et al. (2024). *Tax Flight? Britain’s Wealthiest and Their Attachment to Place*. URL: https://researchonline.lse.ac.uk/id/eprint/121396/1/III_Working_Paper_131_Tax_Flight.pdf (visited on 1 July 2026).
- Leunig, T. (2026). *Fiscal Options: An Argument Against a Wealth Tax*. Nesta. URL: <https://www.nesta.org.uk/blog/fiscal-options-an-argument-against-a-wealth-tax/> (visited on 1 July 2026).
- Neate, R. (2023). *Super-Rich Abandoning Norway at Record Rate as Wealth Tax Rises Slightly*. The Guardian, 10 April. URL: <https://www.theguardian.com/world/2023/apr/10/super-rich-abandoning-norway-at-record-rate-as-wealth-tax-rises-slightly> (visited on 5 July 2026).
- OECD (2018a). *Tax Policy Reforms 2018: OECD and Selected Partner Economies*. Paris.
- (2018b). *The Role and Design of Net Wealth Taxes in the OECD*. 26. Paris.
- Perret, S. (2021). “Why Were Most Wealth Taxes Abandoned and Is This Time Different?” In: *Fiscal Studies* 42.3–4, pp. 539–563.
- Rutter, J. and G. Tetlow (2025). *Wealth Taxes*. Institute for Government. URL: <https://www.instituteforgovernment.org.uk/explainer/wealth-taxes> (visited on 1 July 2026).
- Saez, E. and G. Zucman (2022). “Wealth Taxation: Lessons from History and Recent Developments”. In: *AEA Papers and Proceedings* 112, pp. 58–62.
- Tax Justice UK (2025). *Debunking 5 Common Myths about Wealth Taxes*. Tax Justice UK. URL: <https://taxjustice.uk/blog/debunking-5-common-myths-about-wealth-taxes/> (visited on 1 July 2026).

War, Oil and the Environmental Price of Geopolitics

Aditya Biju

Introduction

A missile that strikes an oil depot does not stop at the border of the country where it lands. In Tehran, attacks on oil and fuel facilities sent smoke over a city of millions, damaged storage tanks, and raised concerns about contamination entering drainage networks. Far from Iran, the same conflict raised the cost of moving fuel, insuring ships and securing fertiliser. What began as destruction of infrastructure became a chain of costs carried by households, farmers, firms, and governments with no direct role in the war. This article argues that the US/Israel-Iran war should be understood as an environmental-economic shock, not only as a geopolitical crisis. The claim focuses on one question: how does a regional war produce environmental damage and transmit that damage through global markets?

The answer is threefold. First, the war creates direct environmental harm through oil fires, damaged military and industrial sites, toxic residues, greenhouse gas emissions, and risks to water and marine environments (UNEP 2026). Secondly, because the Gulf remains a major route for oil, gas and fertiliser trade, these environmental damages translate into price shocks. Finally, the conflict strengthens the economic case for reducing dependence on imported fossil fuels. In Europe especially, the war reinforces an argument already sharpened by Russia's invasion of Ukraine: fossil-fuel dependence is not just a climate risk. It is also a security risk.

An objection to the latter part of this argument is that energy transition cannot be treated as a direct consequence of one war. Europe's move away from fossil fuels began before the Iran conflict and has many causes: climate targets, technology costs, industrial policy and the experience of Russian gas dependence. This objection is fair. The Iran war did not create Europe's energy transition. However, it changed the evidence available to policymakers. It showed again that fossil-fuel dependence exposes economies to military decisions, maritime chokepoints and shocks that no central bank can fully control. The war therefore is not the cause of the transition, but it strengthens the economic case for it.

Direct Environmental Damage

The first channel through which war causes environmental damage is physical destruction. The Conflict and Environment Observatory (CEOBS) identified attacks on Iranian oil

and fuel infrastructure, including the Shahrān fuel and gas depot and the Tehran refinery (CEOBS 2025). The Tehran refinery has a reported capacity of about 225,000 barrels per day, while the Shahrān depot is described as one of Tehran’s largest fuel storage and distribution hubs, with 260 million litres of capacity across 11 storage tanks (CEOBS 2025). When infrastructure of that scale burns near a major city, the consequences extend beyond the immediate target. The environmental risks are severe. Major oil fires release dangerous pollutants such as particulate matter, nitrogen oxides, carbon monoxide, sulphur dioxide, and volatile organic compounds (CEOBS 2025). CEOBS also reported concerns that oil spillages from the Shahrān site entered storm drains in Tehran; a drainage network can move contamination (CEOBS 2026a). In a city of more than nine million people, with geography that can trap air pollution, the health risks are acute (CEOBS 2026a). Residents inhale polluted air before any market indicator captures the event.

Military and industrial sites add another layer of harm. Airfields, missile sites, fuel stores and weapons depots can contain threatening chemicals (oils, heavy metals, energetic compounds, PFAS from firefighting foams, toxic residues from munitions) (CEOBS 2025, 2026b). When such sites burn or explode, some materials are destroyed, but others disperse. The environmental damage then appears as contaminated land, unsafe water, worsened public health, and future remediation costs.

Pollution is not only an environmental loss. It creates costs that are borne by households, health systems, local businesses, and future public budgets. Illness reduces labour productivity, damaged fisheries reduce income, and soil pollution lowers the value of agricultural land. However, none of these costs appear on a military balance sheet. Reconstruction creates a further problem. Destroyed assets must be rebuilt with cement, steel, fuel and machinery, all carbon-intensive materials. If infrastructure is destroyed and then rebuilt, national accounts may record new construction activity as growth. Yet society would only replace what it has lost. This is a weakness of GDP when used as a welfare measure. It can count rebuilding after destruction without asking whether people are better off. Climate accounting faces the same weakness. Civilian firms increasingly measure and report emissions, while conflict-related emissions remain far less transparent. Aircraft fuel, oil fires, emergency logistics, weapons production and reconstruction all add to the carbon burden. If these sources remain poorly measured, war becomes a bigger part of the carbon economy, reducing overall accountability over carbon emissions.



Figure 1: Photo by Yngvi Hegemony on Unsplash

From Local Damage to Global Prices

The second impact runs through global energy and commodity markets. The Strait of Hormuz normally carries 25–30% of the world’s traded oil and about 20% of global LNG

trade, alongside around one-third of global fertiliser shipments (IMF 2026a). These figures explain why even the threat of disruption can move prices. Markets do not wait for a complete closure. Traders price uncertainty, tanker owners demand higher compensation for war risk, and insurers raise premiums (IMF 2026a). Importers look for alternative cargoes, and governments consider using reserves. Each response may be sensible in isolation, but together they raise costs across the economy.

Oil and gas are not ordinary goods. They enter transport, electricity, manufacturing, construction, shipping, and agriculture. Higher oil prices raise diesel and petrol costs, and higher gas prices affect electricity and fertiliser. More expensive transport then raises the cost of moving food from farms to towns and ports to consumers. Airlines, shipping firms and logistics companies pass on costs where they can. A regional security shock therefore causes cost-push inflation, creating a difficult problem for central banks. If inflation comes from domestic demand, higher interest rates can cool spending. If it comes from a war-related fuel shock, monetary tightening cannot reopen a sea lane or put out an oil fire. It can only reduce demand elsewhere in the economy, meaning such shocks carry stagflationary risk: prices rise while growth weakens.

Furthermore, the burden falls unevenly. Some commodity exporters with large reserves and fiscal buffers, such as Saudi Arabia and the UAE, can absorb market stress more easily. They may even gain revenue when prices rise, although they remain exposed to security risks and higher risk premia. Importers face the opposite problem: rising fuel bills. Consequently, their trade balances may worsen, and their currencies can come under pressure. For poorer economies, the effect is sharper because food and energy take a larger share of household income and public budgets (IMF 2026a). India, Bangladesh and Pakistan illustrate the vulnerability of Asian importers. The International Energy Agency (IEA) notes that these three countries imported almost two-thirds of their LNG supplies through the Strait of Hormuz in 2025 (International Energy Agency 2026). A disruption to Qatari and Emirati LNG would therefore tighten energy supply in markets where affordability already matters acutely. In Bangladesh and Pakistan, where gas contributes heavily to electricity generation, reduced LNG availability could weaken power security and affect gas-intensive industries, including fertiliser production. The point is not that every restaurant closes, or every household loses cooking fuel. It is that a maritime shock can move quickly into electricity, fertiliser and food systems.

The fertiliser market makes the transmission clearer. Natural gas is a major input in ammonia production, which supports nitrogen fertilisers. When gas prices rise or fertiliser shipments are disrupted, farmers face higher costs. Some reduce application, risking lower yields. Others maintain usage and pass costs through higher food prices. The World Bank projected fertiliser prices to rise by more than 30% in 2026, after disruptions connected to the Strait of Hormuz (World Bank 2026). The effect does not appear only at the moment of conflict. It can emerge months later through planting decisions, harvests, and retail food prices. Obviously, not every food-price increase after the war can be attributed to the conflict. Weather, domestic policy, exchange rates, and local supply chains also matter. Still, the mechanism is clear. Fossil-fuel shocks do not stay in the energy sector, but enter food systems through transport, fertiliser, and currency pressure.

Europe and Energy Security

The indirect effect of the conflict is most clearly illustrated in Europe. After Russia's invasion of Ukraine, the European Union treated energy dependence as a strategic weakness. REPowerEU aimed to reduce reliance on Russian fossil fuels and accelerate clean-energy deployment (European Commission 2024; European Investment Bank 2026). The Iran war further strengthens this argument's foundation. Even if Europe removes Russian gas from its energy mix, fossil-fuel dependence can still expose it to shocks from the Gulf, the Red Sea and other chokepoints. Investment in low-carbon energy therefore has a security rationale, alongside its climate rationale. One could argue that solar and wind power do not eliminate all risks. They require grids, storage, flexible demand and planning reform. Nuclear power is expensive, slow to build and politically contested. It also requires strong regulation and public trust. Nonetheless, these technologies reduce dependence on daily flows of imported fossil fuels, which are subject to global geopolitical events.

The economic comparison has therefore changed. The question is not simply whether fossil fuels are cheaper than clean energy in a calm year. The better question is whether an energy system remains affordable once price volatility, war risk, carbon damage and strategic dependence are accounted for. While a gas-fired power station may appear cheaper under normal market conditions, it looks less cheap if its fuel price can skyrocket due to a distant war.



Figure 2: Photo by Markus Spiske on Unsplash

Of course, reducing import dependence does not necessarily require abandoning fossil fuels. Countries with domestic reserves can instead increase home-grown production, at least during the transition. The UK provides a useful example. Expanding North Sea oil and gas has become a contentious fault line in British politics: Nigel Farage and Reform UK have advocated substantially greater extraction, while Prime Minister Andy Burnham has adopted a more cautious but increasingly “pragmatic” position, arguing that domestic resources cannot simply be ignored while households and industry face high energy costs (Reuters 2026).

The economic case is not trivial. Domestic production can reduce import volumes, preserve employment and tax revenues, and provide some protection against physical supply disruption. Yet it offers far less protection against the price shock at the centre of this article. North Sea oil and gas are sold into international markets, so British consumers remain exposed to global prices even when the fuel originates domestically. The North Sea basin is also mature: production has fallen by more than 70% since its 1999 peak, leaving the UK a net importer of both oil and gas (Hewitt and Bolton 2026). Increasing domestic extraction may therefore soften import dependence at the margin, but it cannot insulate an economy from a globally traded fossil-fuel market. In that respect, it is a complement to short-term energy security rather than an alternative to longer-term diversification.

This distinction helps explain the broader shift in energy investment. In 2025, global energy investment was expected to reach about USD 3.3 trillion, with roughly USD 2.2 trillion going to renewables, nuclear, grids, storage, low-emissions fuels, efficiency and electrification (International Energy Agency 2025). This sum was about twice the amount going to oil, gas and coal, showing that capital allocation has already moved towards electricity-based and low-carbon infrastructure (International Energy Agency 2025). Public investment has a strong economic justification here. Private firms may underinvest in resilience because they cannot capture all its benefits: a stronger electricity grid helps households, hospitals, factories and public services. Furthermore, storage reduces exposure to fuel-price spikes, and domestic nuclear and renewable energy sources reduce dependence on insecure supply routes. These benefits spread across society, so governments have reason to invest even where private returns look uncertain.

The strongest counterargument is that clean energy can create new dependencies. Solar panels, batteries, critical minerals, uranium supply chains and grid equipment all involve geopolitical risks. This statement is true; an energy transition does not abolish dependence, but rather changes its form. A poorly planned transition could replace dependence on Gulf oil with dependence on concentrated mineral processing or imported equipment. Therefore, the policy goal should not be a single technology or a baseless slogan. It should be a diversified energy portfolio with domestic capacity, storage, interconnections, strategic reserves, recycling and resilient supply chains.

Conclusion

The Iran war produces a harsh short-term result. It increases pollution, damages infrastructure, raises emissions and creates risks to water, soil and public health. It also raises costs for countries far from the battlefield, through the impacts on the Strait of Hormuz. These costs are not distributed fairly. Poor households and import-dependent economies face the sharpest pressure. Over a longer period, the same conflict may strengthen the political economy of energy transition. Climate policy often struggles because some benefits appear distant. Energy security makes the benefit immediate. A country that builds cleaner domestic power, stronger grids and lower fuel demand reduces exposure to current and future shocks as well as emissions.

This is not an argument that war is useful for climate policy. That conclusion would be wrong. If governments need conflict to recognise the risks of fossil-fuel dependence, then policy has already failed. The climate case, public-health case and security case for reducing reliance on traded hydrocarbons existed before the war. The conflict has made the cost of delay clearer.

The lesson is therefore practical rather than rhetorical. The US/Israel-Iran war shows that fossil-fuel dependence globalises conflict. A strike near Tehran can raise costs in Europe. A threatened sea lane can affect fertiliser prices in Asia. A burning refinery can become a local health crisis and a global carbon burden. Energy transition should be treated as economic insurance as well as environmental policy. Renewables, nuclear power, grids, storage and efficiency will not abolish geopolitical risk. They can reduce the extent to which distant wars enter domestic economies through fuel prices, food costs and inflation. The smoke rises locally, but the economics is global. A serious energy

policy must begin from that fact.

References

- CEOBS (2025). *The Emerging Environmental Consequences of the Israel-Iran War*. Conflict and Environment Observatory. URL: <https://ceobs.org/the-emerging-environmental-consequences-of-the-israel-iran-war/> (visited on 15 July 2026).
- CEOBS (2026a). *Black Rain: The Health and Environmental Risks from Tehran's Oil Fires*. Conflict and Environment Observatory. URL: <https://ceobs.org/black-rain-the-health-and-environmental-risks-from-tehrans-oil-fires/> (visited on 15 July 2026).
- CEOBS (2026b). *Operation Epic Fury: Emerging Environmental Harm and Risks in Iran and the Region*. Conflict and Environment Observatory. URL: <https://ceobs.org/operation-epic-fury-emerging-environmental-harm-and-risks-in-iran-and-the-region/> (visited on 14 July 2026).
- European Commission (2024). *Two Years of REPowerEU: Strengthening Europe's Energy Resilience*. European Commission. URL: https://commission.europa.eu/news-and-media/news/two-years-repowereu-strengthening-europes-energy-resilience-2024-05-17_en (visited on 24 July 2026).
- European Investment Bank (2026). *REPowerEU and the EIB*. European Investment Bank. URL: <https://www.eib.org/en/projects/topics/energy-natural-resources/energy/repowereu> (visited on 24 July 2026).
- Hewitt, T. and P. Bolton (2026). *Debate on North Sea Oil and Gas*. Research Briefing CBP-10905. House of Commons Library, 23 June. URL: <https://researchbriefings.files.parliament.uk/documents/CBP-10905/CBP-10905.pdf> (visited on 14 September 2026).
- IMF (2026a). *How the War in the Middle East Is Affecting Energy, Trade, and Finance*. International Monetary Fund. URL: <https://www.imf.org/en/blogs/articles/2026/03/30/how-the-war-in-the-middle-east-is-affecting-energy-trade-and-finance> (visited on 17 July 2026).
- International Energy Agency (2025). *World Energy Investment 2025: Executive Summary*. International Energy Agency. URL: <https://www.iea.org/reports/world-energy-investment-2025/executive-summary> (visited on 18 July 2026).
- International Energy Agency (2026). *Strait of Hormuz Factsheet*. International Energy Agency. URL: <https://www.iea.org/about/oil-security-and-emergency-response/strait-of-hormuz> (visited on 14 September 2026).
- Reuters (2026). *UK PM Burnham Says He Will Use North Sea Oil Pragmatically After Trump Comments*. Reuters. URL: <https://www.reuters.com/world/uk/uk-pm-burnham-says-he-will-use-north-sea-oil-pragmatically-after-trump-comments-2026-07-30/> (visited on 14 September 2026).
- UNEP (2026). *UNEP Statement on Environmental Damage Arising from the Conflict in the Middle East*. United Nations Environment Programme. URL: <https://www.unep.org/news-and-stories/statements/unep-statement->

environmental-damage-arising-conflict-middle-east (visited on 12 July 2026).

World Bank (2026). *Middle East War to Spark Biggest Energy Price Surge in Four Years*. World Bank. URL: <https://www.worldbank.org/en/news/press-release/2026/04/28/commodity-markets-outlook-april-2026-press-release> (visited on 14 September 2026).

Contracts No Regulator Has Graded: The Information Question America Answered Blind and Britain Never Asked

Dan Low

In June 2026, \$44.8bn changed hands on Kalshi and Polymarket (Park 2026). The owner of the New York Stock Exchange committed up to \$2bn to Polymarket in October 2025, and Kalshi is reported to be seeking new investment at a \$40bn valuation (ICE 2025; Braun 2026). Rules governing what these exchanges may list are being written now in the United States and in Britain, and neither has measured how well these contracts forecast anything.

The author's own unpublished study grades the contracts those exchanges list against their own prices. The short version is that they are very close to perfectly calibrated: across 1,206,483 settled contracts on financial and macroeconomic outcomes, the prices printed between 70 and 75 cents were followed by the event 72.79 per cent of the time. Miscalibration costs them 0.07 per cent of what they could in principle achieve.

That honesty turns out to be the cheap half. What separates a useful forecast from a merely honest one is resolution: whether the cases a market prices differently do in fact turn out differently. The skill available to a contract is the most any price could have extracted from a question that hard, given how often contracts like it come true at all, so the share captured is comparable across subjects that are not equally hard. Across the sample these contracts capture 50.4 per cent of it, and almost none of what they leave behind is lost to miscalibration. It is lost to resolution: what they capture varies by more than a factor of two across what a contract is about, and the shortest contracts capture the least of what is available to them.

The distinction is not a technicality here. The American rulemaking asks a question only this axis answers, and the British prohibition never asked it.

The rules being written

The CFTC's proposed rule of June 2026 turns on whether a contract involves an enumerated activity (unlawful conduct, terrorism, assassination, war, gaming, or other similar activity the Commission designates by rule or regulation) and whether it is contrary to the public interest (CFTC 2026a). The litigation has been fiercer than the rulemaking, and almost all of it is about sport. In April a federal appeals court held that sports event contracts are likely swaps under federal law and that federal regulation displaces New

Jersey’s gambling laws, upholding an injunction against the state’s gaming regulator by two votes to one (*KalshiEX v. Flaherty* 2026). In July a judge in Manhattan reached the opposite result on New York’s law, finding Kalshi unlikely to succeed on the argument that federal law displaces New York’s gambling statutes and assuming without deciding that the contracts are swaps at all (*KalshiEX v. Williams* 2026). In August the Ninth Circuit split the courts of appeals, holding that these contracts are likely not swaps and saying so explicitly against the Third (*KalshiEX v. Assad* 2026). Three more circuits are weighing it, the Commission has sued eight states and intervened against a ninth to assert its jurisdiction (CFTC 2026b), and the question is headed for the Supreme Court.

The financial and macroeconomic contracts studied here take a quieter route. Exchanges list them by self-certification, which means filing a contract’s terms with the Commission, attesting that those terms comply, and listing a day later unless it intervenes. No approval is issued and none is required. The economic-purpose argument does the work: the exchange’s own case that the contract is tied to something of economic consequence. Crypto.com’s derivatives arm certified a Bitcoin event contract in April on the grounds that it expresses a view on the price of Bitcoin, an asset whose price carries broad economic and commercial consequences (CDNA 2026). A separate CFTC proposal on data reporting for event contracts closed for comment in July (CFTC 2026c).

None of this is unregulated. It is lightly examined, and the standard it will be examined against is a question about information rather than about risk. The Commission’s proposed rule asks whether a contract provides “meaningful hedging or price-basing utility”, yields “economically useful or otherwise meaningful information”, or promotes responsible innovation and fair competition. The three limbs are disjunctive, and on the first the Commission is explicit that it “is not necessary to demonstrate that event contracts have a reasonable potential for a hedging or pricing function” to escape an adverse finding, though such potential would count in a contract’s favour (CFTC 2026a).

The Commission adds that a contract’s price discovery need not stand alone. It can arise, the rule says, from how the prices of a variety of event contracts can be factored into decisions, not just from whether any one of them is useful by itself. That is the level at which the evidence below operates.

That question about information is one the study’s data answers, and it needs both halves of the split. A price that never separates one case from another has aggregated nothing; a price that separates them and labels them wrongly is information nobody can use. These markets have the labelling (calibration); what varies is the separation (resolution). By that standard the listings are lopsided. Contracts on central bank decisions capture 86.7 per cent of the skill available to them, the highest of any category here; the crypto contracts manage 49.4 per cent. The same study finds those central bank contracts the worst calibrated corner of the sample. Both things are true of one set of prices, which is the point: the honesty and the separation are earned separately. Some of the gap between the central bank contracts and the crypto ones belongs to the question rather than to the market: a rate decision becomes knowable well before it is announced, and a contract that spends weeks parked near certainty earns its resolution cheaply. The gap survives the most conservative intervals this data supports, and the agency’s test in any case asks what a contract tells you, not how hard it worked to find out.

That contrast does not generalise. Single stocks and foreign exchange also capture more than crypto, whilst inflation and labour-market contracts capture less. The ordering is not macroeconomic contracts against the rest, and it does not run with how much a category is traded. By contract length the ends of the range do separate. Five-minute contracts capture 30.9 per cent of the skill available to them against 59.8 per cent for weekly ones, and the graded contracts crowd into the short end: 759,112 run for an hour and 119,540 for five minutes, against 7,534 monthly. The middle of that range does not climb steadily with length, though: hourly contracts capture more than four-hourly or daily ones, and the share peaks at the weekly contracts before falling back for monthly ones, which is not what one would expect if a contract's resolution were only a matter of how early its answer becomes knowable. Macroeconomic contracts number 9,935 in all, well under one per cent of the contracts graded here. The instruments that best satisfy a test about information are a rounding error. The ones that dominate are neither the best nor the worst: the hourly contracts capture 47.6 per cent, a little under the sample's own 50.4 per cent.

The agency has not measured any of this. Its proposed rule says these prices may reflect outcomes better than survey-based forecasts, and the words it puts in quotation marks come from a Federal Reserve working paper built on about a dozen Kalshi series (CFTC 2026a; Diercks et al. 2026). Its guide for the public drops the hedge and the citation both, and simply tells readers that prediction markets can sometimes forecast event outcomes better than polling (CFTC n.d.).

It has not lacked the time. The question was put, and put well, in 2008: a Commission concept release on event contracts asked what calculations, analyses and factors could be used to determine the social value of the information such trading discovers. The same release observed that the Iowa Electronic Markets appeared to have produced better predictive indicators than professional polling, and footnoted a law review article rather than any work of its own (CFTC 2008). Twenty-four numbered questions went out. No rule followed. The file reopened in March 2026 with an advance notice that posed forty fresh questions of its own, none of them about accuracy (CFTC 2026d); comments closed that April and the proposal arrived in June. The proposal's own view is that effective price discovery relies on robust metrics such as forecast accuracy, a belief it states preliminarily and tests nowhere (CFTC 2026a). Eighteen years separate the question from that sentence, and in between the agency measured nothing.

The courts have not asked either, and one of them came close to having to. When Kalshi sued the Commission over its congressional-control contracts, the district court held that "gaming" means playing a game for stakes, that an election is not a game, and that the public-interest review was therefore never triggered at all; the Commission dropped its appeal in May 2025 (*KalshiEX v. CFTC* 2024). The economics went untested. The Commission has twice rejected an economic-purpose case on its merits, against Nadex's election contracts in 2012 and against Kalshi's congressional contracts in 2023, and both times the ground was that the link between the event and any real exposure was too remote to hedge. Nadex is now Crypto.com's derivatives arm, so the registrant refused on economic-purpose grounds in 2012 is the one self-certifying in 2026. The 2012 order went further and found the contracts had no price-basing utility, which is a finding about information; the present proposal says that reasoning was incorrect (CFTC 2026a).

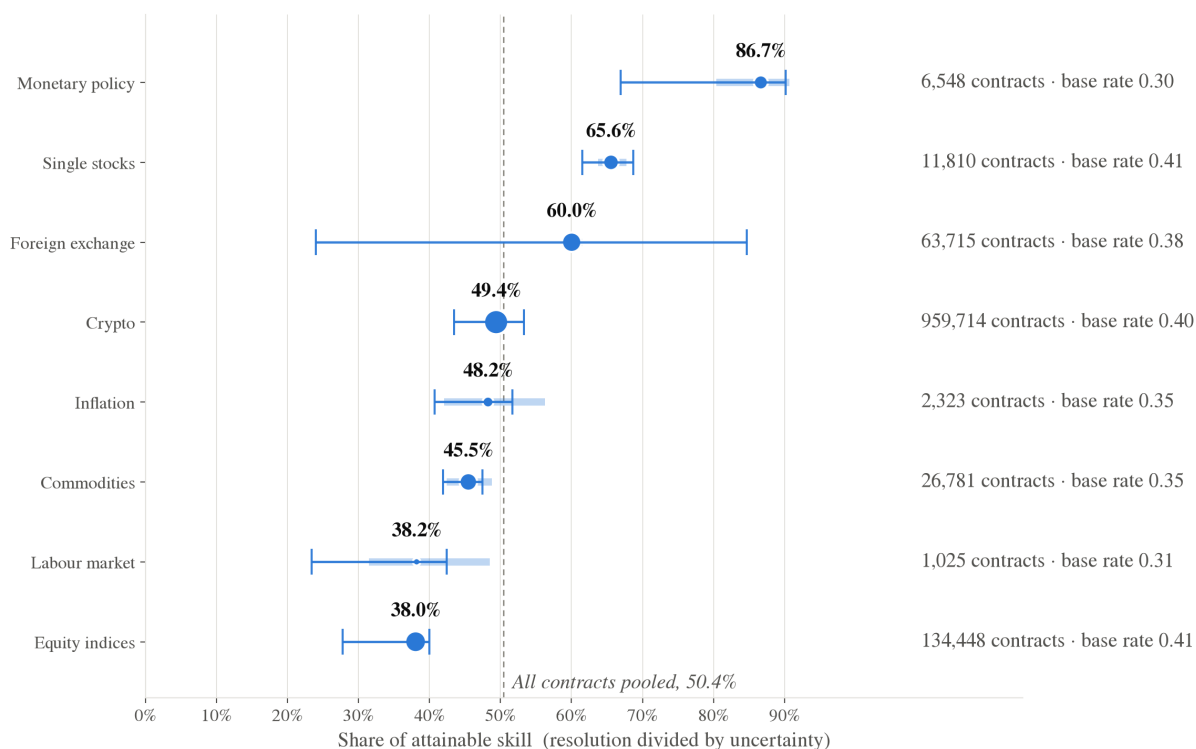


Figure 1: *Share of attainable skill, by what a contract is about.* Resolution divided by uncertainty, so that categories with different base rates can be compared. Thick pale bars are 95 per cent intervals clustered by contract and thin dark lines by contract family, usually the more conservative of the two, though on inflation and commodities the contract-level bar is the wider and on four of the eight rows the two cross rather than nest. On foreign exchange, crypto and equity indices the contract-level bar is narrower than the dot and sits hidden behind it, so those three rows show a family-level line only. Dot area scales with the log of the contract count, which is printed beside each row. The dashed line is the pooled figure for the sample as a whole. Monetary policy and crypto do not overlap even at the coarser clustering, which is what answers the objection that a very thin cell is being set against a very thick one. The two largest categories both sit at or below the pooled average, and the ordering does not otherwise run with size. Categories of fewer than 1,000 contracts are not shown. Counts are of graded contracts in the calibration sample, not of everything listed.

The gap has not gone unremarked from inside. Kristin Johnson, leaving the Commission in September 2025, named the problem as “too few guardrails and too little visibility into the prediction market landscape” (Johnson 2025). The accuracy claim has been put to the Commission from outside the building too, and put early: four academics who work on these markets told it in 2022 that prediction market forecasts “outperform other key predictions in: accuracy, latency, and time-granularity” (Crane et al. 2022). That may well be true on average. What this data shows is that it may not be true uniformly, and the variation is the whole question.

What it costs to be told

One further thing is measurable here, and it is a cost rather than a quality of the forecast.

Many events are listed as ladders: mutually exclusive contracts covering a range of outcomes. Where a ladder is exhaustive, exactly one member comes true, so the prices should sum to a dollar. Anything above that is what it costs to trade the whole thing, the quantity a bookmaker calls the overround. It is the only cost-like measure available in this data, which records executed trades but no order book, and so no quoted spread.

On ladders that genuinely span every outcome, the excess over a dollar is 13.0 percentage points on Kalshi against 0.9 percentage points on Polymarket. Such ladders are 4.7 per cent of Kalshi’s and 90.7 per cent of Polymarket’s.

Kalshi charges a taker fee on fills and Polymarket charged no trading fee at all until January 2026 (Kalshi 2026; Polymarket 2026); on the exhaustive ladders behind this comparison at least 89.5 per cent of Polymarket’s charged nothing, so the fee is the obvious candidate for the gap. Note 1 works out that it is at most about half an answer: the fee vanishes as a contract’s price approaches either end, and passed into the printed prices in full it could not add more than seven percentage points to a ladder’s sum, against the 13.0 percentage points measured above. The method note sets out what these data can and cannot say about the fee’s contribution.

The reading to resist is that the rest is simply Kalshi charging more. Its exhaustive ladders trade far less than Polymarket’s, and the least traded ladders carry the largest excesses on both exchanges, so liquidity is the obvious thing to control for. It cannot be held fixed here, because liquidity and date move together: hold the comparison to the window in which both exchanges list such ladders and the gap widens rather than narrows. Kalshi’s own figure also moves too much from one year to the next to read as a price list. One control does cut the other way, and it is in the method note: tightening the staleness cap from eight hours to fifteen minutes, so that both exchanges are held to prices struck close together, roughly halves the gap without closing it, though not steadily. That measurement is made across all complete ladder snapshots rather than the exhaustive ones, where the sweep has not been run. What is left cannot be split into spread against inventory, because the data records executed trades and no quotes.

You cannot buy any of this

British readers can follow this argument but not join it, for reasons that invert the American ones.

British law splits event contracts by what they reference. Contracts on sport or politics fall to the Gambling Commission, which said in February 2026 that current prediction market products would appear to fall within the definition of a betting intermediary, much as an established betting exchange does (Gambling Commission 2026). Contracts referencing financial instruments, economic statistics, commodities or interest rates go the other way, because a bet that is already a regulated activity under the Financial Services and Markets Act is carved out of the Gambling Act's definition of betting (Gambling Act 2005, s.10). That carve-out moves with the financial perimeter rather than with gambling law: an order under section 22 can bring a class of bet inside the Financial Services and Markets Act or take it out again, and the Gambling Act follows automatically, so the boundary can be redrawn without amending either statute. The FCA banned the sale of binary options to retail consumers in 2019 on grounds of product risk, firm conduct and consumer harm; announcing it, Christopher Woolard called them "gambling products dressed up as financial instruments" (FCA 2019).

The FCA's own perimeter report, updated in July 2026, lists prediction market products as an outside-the-perimeter issue and flags further work on access and on clarifying where the boundary falls (FCA 2026). Asked in the Lords in February what assessment the government had made of these products' risks and costs, the minister answered in licensing terms and undertook to monitor them and consider further action if necessary (Twycross 2026): a question about evidence, answered with a promise to look. The boundary itself is not in doubt, and the regulator has said where it falls. What nobody has said is whether the contracts on the financial side of it are any good at the thing they exist to do.

The contracts studied here are financial ones. They fall, in other words, into the category the FCA's 2019 ban covers. Neither exchange is open to British users: Kalshi's member agreement prohibits them (Kalshi n.d.), and Polymarket blocks British addresses (Polymarket n.d.).

Where a line would fall

Both jurisdictions have drawn a line, and neither line falls where this evidence would put it. The United States is arguing over whether sports contracts are derivatives or bets. Britain decided the financial ones were the dangerous category and prohibited them for retail consumers. The Commission has answered the information question without ever measuring the information; Britain has not put the question at all. Which contracts actually know anything is answerable, and this data answers it.

What such a test would look like is not mysterious. The Commission already proposes to collect the data it would need: its July notice, otherwise a relief measure, would require transaction reporting on these contracts (CFTC 2026c). Grading a contract family against its own settlements is arithmetic after that. A rule could ask an exchange

to show what share of the available skill the family a proposed contract belongs to has historically captured, and leave the Commission to set the threshold. Across the subjects graded here that share is 86.7 per cent on central bank decisions against 49.4 per cent on the crypto contracts that dominate the listings; across contract lengths it runs from 30.9 per cent to 59.8 per cent. Where that line falls would decide something. That would be a more demanding standard than the one now on the table. It would also be a more honest one, because it is a question with an answer.

Method

The figures quoted here are drawn from the author’s unpublished study: a frozen data cut, 18 August 2026, of 1,206,483 settled contracts and 32,092,831 traded minutes, from which reliability, resolution and skill share are computed by Murphy decomposition (Murphy 1973). Reliability across the sample is 0.00017 against an uncertainty of 0.2384, and their ratio is the 0.07 per cent quoted at the top. The band between 70 and 75 cents holds 784,654 traded minutes across 216,697 contracts, so a contract enters it once for every minute in which it traded. Three of the study’s results are set out below: the share of attainable skill by asset category, the same share by contract length, and the overround. The skill shares come from the whole of that sample. The overround comes from a much narrower population: 18,201 ladders, taken as 19,266 snapshots of a whole ladder’s prices at a single moment, because prices can only be added up across contracts that are mutually exclusive. A snapshot is one reading of one ladder at one of five fixed distances before its fixing: a week, a day, four hours, an hour, and fifteen minutes. A reading counts only where every contract in the ladder carries a price struck recently enough for that distance, and most ladders clear that bar at a single distance, which is why the two counts sit so close together. Polymarket’s ladders, priced more continuously, clear it at more distances than Kalshi’s.

Skill share across asset categories, for the comparison drawn in the text, with the number of graded contracts behind each: Monetary policy 86.7 per cent ($n = 6,548$); Single stocks 65.6 per cent ($n = 11,810$); Foreign exchange 60.0 per cent ($n = 63,715$); Crypto 49.4 per cent ($n = 959,714$); Inflation 48.2 per cent ($n = 2,323$); Commodities 45.5 per cent ($n = 26,781$); Labour market 38.2 per cent ($n = 1,025$); Equity indices 38.0 per cent ($n = 134,448$). Two further categories are excluded from both Figure 1 and the list above, for carrying fewer than 1,000 graded contracts each: ETFs ($n = 80$), GDP ($n = 39$).

Skill share by contract length, for the comparison drawn in the text, with the number of graded contracts behind each: weekly 59.8 per cent ($n = 39,228$); monthly 49.5 per cent ($n = 7,534$); hourly 47.6 per cent ($n = 759,112$); four-hourly 42.8 per cent ($n = 12,646$); daily 37.4 per cent ($n = 60,832$); fifteen-minute 34.5 per cent ($n = 205,517$); five-minute 30.9 per cent ($n = 119,540$). Identified contract lengths carrying fewer than 1,000 graded contracts are not listed: annual ($n = 341$), quarterly ($n = 39$). Contracts whose length is not identified are excluded from this stratification rather than listed as a length of their own.

On the overround, its benchmark and the controls applied to it. On a set of contracts that are mutually exclusive and pay a dollar between them, the member prices should

sum to the present value of a dollar rather than to a dollar. They sum instead to a median of 1.120 on Kalshi and 1.009 on Polymarket, an excess over a dollar of 12.0 and 0.9 percentage points. Discounting cannot account for any of that, and not because it is too small: it pushes the other way. A dollar paid at settlement is worth less than a dollar now, so the true benchmark sits below a dollar and the real excess is larger still, by the discount factor over the remaining term, roughly 40 basis points on a monthly ladder and nothing worth measuring on an hourly one.

Most Kalshi and some Polymarket ladders are bounded bands with no open-ended rung at either end, so they are not exhaustive and their no-arbitrage value is the chance of landing inside the band rather than the whole dollar, empirically 97.1 per cent against Polymarket's 98.7 per cent. Measured against that benchmark the overruns across all ladder snapshots are 15.4 and 2.3 per cent of the fair price. Restricted to ladders that genuinely do span every outcome, where a dollar is the right benchmark and no outcome data is needed at all, the medians are 1.130 on 902 Kalshi snapshots and 1.009 on 1,374 Polymarket ones.

That comparison is not like for like, and the reason is the calendar. Kalshi's exhaustive ladders run from 4 May 2022 to 21 July 2026 and Polymarket's only from 31 October 2024 to 27 April 2026, so the pooled figure sets most of Kalshi's history against an exchange that was not yet listing this kind of ladder. Kalshi's own figure is strongly time-varying, which is the more important fact. By year of snapshot it runs: 2022, 9.0 percentage points on 285 snapshots; 2023, 5.0 on 99; 2024, 8.0 on 139; 2025, 115.0 on 5; and 2026, 45.5 on 374. A single pooled median over a quantity that moves this much describes none of the years it averages: from its first year to its last it multiplies by 5.1, and from its narrowest year to its widest by 23, the widest resting on 5 snapshots and the narrowest on 99. A cost of trading does not behave like that.

Under an explicit window the gap widens rather than narrows. Over the window where both exchanges have bars, 12 December 2022 to 28 April 2026, Kalshi is 1.200 on 589 snapshots against Polymarket's 1.009 on 1,374, a gap of 19.1 percentage points. Over the window where both list exhaustive ladders, 31 October 2024 to 27 April 2026, Kalshi is 1.510 on 352 snapshots against the same 1,374, a gap of 50.1 percentage points. Neither window is preferred here. The first is the wider of the two and inherits the calendar problem set out above; the second is the only window in which both exchanges are genuinely present, and it is also the one that discards 550 of Kalshi's exhaustive snapshots. That second window is also very nearly a single calendar year: 339 of its 352 Kalshi snapshots fall in 2026, 96 per cent of them. The control that widens the gap therefore rests on one year of the series set out above rather than on the whole of Kalshi's span.

Liquidity can be measured here but not held fixed. With the ladders sorted into bands by the median traded minutes per contract, the thinnest band holds 497 of Kalshi's exhaustive snapshots against 17 of Polymarket's, and it carries the largest excess on both exchanges, with sums of 1.290 and 1.014 dollars. The bands match on liquidity and not on date, though, and once the comparison is also held to the window where both exchanges list exhaustive ladders, no band keeps enough ladders on both sides to compare at all. Each band's own figure stands as a measurement; a single liquidity-adjusted gap across all of them does not, because there is no band left in which to compute one. One control does survive, and it survives in sign rather than in size: Kalshi's median stays above

Polymarket’s at all six staleness caps tested, so the gap is not created by prices struck far apart. That sweep is taken across all complete ladder snapshots, 32,264 on Kalshi against 5,701 on Polymarket, rather than the exhaustive ones the rest of this note is about, where it has not been run. Its size is another matter, and the sweep does not decide it. Across the six caps the gap runs from 12.1 percentage points at one hour down to 5.3 at fifteen minutes, and it does not fall steadily as the cap tightens: it is wider at one hour than at eight hours. Only the tightest caps bring it down, so how large a fall the control shows depends on which cell is read as the baseline: 61 per cent from no cap, 51 per cent from eight hours.

What the gap cannot be decomposed into is fee against spread against inventory. Kalshi’s taker fee is a candidate for part of it, but the published schedule bounds it at most at about half the gap, and Polymarket charged no trading fee at all until January 2026. The exchanges’ own fee flag puts a floor under that, in both of the units this comparison uses: at least 91.6 per cent of Polymarket’s exhaustive snapshots sit on a ladder carrying no fee-enabled contract, and at least 89.5 per cent of the exhaustive ladders those snapshots are readings of carry none. Every snapshot on a ladder that carries one falls in 2026: 2024, 0 of 7 (0.0 per cent); 2025, 0 of 860 (0.0 per cent); 2026, 115 of 507 (22.7 per cent). The flag records whether a schedule applies rather than what was charged, so it bounds the fee-free share from below. On Kalshi the flag is set on every exhaustive snapshot in the sample, and that hundred per cent is a ceiling on fee-charging rather than a count of fees actually charged. What is left cannot be split further, because the dataset contains executed fills and no order book at all. That is the limit of what these data can establish.

The figure is Figure 1 above.

The code and the data cut behind these figures are held back pending separate publication of the study they come from.

Notes

1. *“could not add more than seven percentage points”*. Kalshi charges a taker seven per cent of the number of contracts filled, multiplied by the price and again by one minus the price, rounded up so that the fee and the position’s cost together land on a hundredth of a cent (Kalshi 2026). Let p be what a rung would print at if no fee were charged, so that the rungs of an exhaustive ladder sum to a dollar, and suppose the whole of each rung’s fee is passed into the price that rung prints at. That is the most generous assumption available to the fee, and it is the one taken here. The printed prices then sum to $1 + 0.07 \sum p(1 - p)$, which is $1 + 0.07(1 - \sum p^2)$, the sums running over the rungs. $\sum p^2$ is smallest when the rungs are priced alike, at one over their number, so the fee can add at most $0.07(1 - 1/n)$ on a ladder of n rungs: 6.3 percentage points at ten rungs and 6.5 at fourteen, the range the median Kalshi ladder behind the figure above falls in, and never a full seven percentage points however many rungs a ladder carries. The rounding is to a hundredth of a cent on each order, so across a ladder of this size it adds at most about a seventh of a percentage point. The coarser rounding to the next cent, in the schedule Kalshi self-certified in September 2022, is not the rule in force. The bound is not affected by which of the two windows the comparison is drawn over, because it is arithmetic on a ladder rather than a fact about the sample. It does depend on the

fee schedule, which an exchange can change: seven per cent is the general taker rate in Kalshi’s current schedule and in the fee schedule it self-certified with the Commission in September 2022, and no series in either charges a taker more.

References

- Braun, H. (2026). “Kalshi Targets a Massive \$40 Billion Valuation, Widening Lead Over Rival Polymarket”. In: *CoinDesk*, 24 June.
- CDNA (2026). *Product Certification: Bitcoin Event Contract – Submission Pursuant to Commission Regulation 40.2(a)*. Indexed by the Commission as “CDNA Certification of Bitcoin Event Contract – 14.71”. Filed as North American Derivatives Exchange Inc., trading as Crypto.com Derivatives North America. Washington, DC: Commodity Futures Trading Commission, 1 April.
- CFTC (n.d.). *Understanding Prediction Markets and Event Contracts*. Consumer education page, cftc.gov. Tells readers that “when everyone’s predictions and knowledge are combined, prediction markets can sometimes forecast event outcomes better than polling or other forms of forecasting”, with no citation anywhere in that section (visited on 1 September 2026).
- CFTC (2008). *Concept Release on the Appropriate Regulatory Treatment of Event Contracts*. Request for public comment. 73 Fed. Reg. 25669, 7 May. Posed twenty-four questions, including one on the social value of information discovered through event-contract trading. No rule followed.
- CFTC (2026a). *Prediction Markets; Public Interest Determinations*. Notice of proposed rulemaking. 91 Fed. Reg. 35806, 12 June. Docket CFTC-2026-1189, RIN 3038-AF65. Comment period closed 27 July 2026. The proposal arising from the advance notice below.
- CFTC (2026b). *Complaints and Press Releases in the State Actions*. Press releases 9206-26 (Arizona, Connecticut and Illinois, 2 April), 9218-26 (New York, 24 April), 9220-26 (Wisconsin, 28 April), 9233-26 (Minnesota, 19 May) and 9238-26 (Rhode Island, 28 May, an intervention rather than a fresh suit), with further filings against New Mexico on 12 June (9251-26) and Kentucky on 23 June (9260-26). Eight suits and one intervention, alongside amicus briefs in the Massachusetts Supreme Judicial Court (9219-26, 24 April) and the Sixth Circuit (9230-26, 12 May). Complaints brought jointly by the United States and the Commission; the defendants are states and state officials (visited on 1 September 2026).
- CFTC (2026c). *Data Reporting Requirements for Certain Event Contracts*. Notice of proposed rulemaking. 91 Fed. Reg. 40102, 1 July. Comment period closed 31 July 2026.
- CFTC (2026d). *Prediction Markets*. Advance notice of proposed rulemaking; request for comments. 91 Fed. Reg. 12516, 16 March. Docket CFTC-2026-0331, RIN 3038-AF65. Comment period closed 30 April 2026. Became the proposal above.
- Crane, H., D. Pennock, D. Rothschild, and K. Strumpf (2022). *Comment Letter on Prediction Markets*. Washington, DC: Commodity Futures Trading Commission. Comment 69735, filed by Rothschild, on Industry Filing 22-002, the Regulation 40.11 review of KalshiEX’s congressional control contracts, comments.cftc.gov, 22 September. Undated in the letter itself; dated in the Commission’s record (visited on 4 September 2026).
- Diercks, A. M., J. D. Katz, and J. H. Wright (2026). *Kalshi and the Rise of Macro*

- Markets*. Finance and Economics Discussion Series 2026-010. Washington, DC: Board of Governors of the Federal Reserve System, February. Also issued as NBER Working Paper 34702, January 2026.
- FCA (2019). *Product Intervention Measures for Retail Binary Options*. Policy Statement PS19/11. London: Financial Conduct Authority, 29 March. In force from 2 April 2019. The accompanying statement of the same date carries Christopher Woolard, then Executive Director of Strategy and Competition, describing binary options as “gambling products dressed up as financial instruments”.
- FCA (2026). *Perimeter Report*. London: Financial Conduct Authority, first published 26 March, updated 16 July (visited on 1 September 2026).
- Gambling Act 2005*. s.10 (“Spread bets, &c.”). For the purposes of s.9(1), a bet the making or accepting of which is a regulated activity within the meaning of s.22 of the Financial Services and Markets Act 2000 is not a bet. The carve-out is defeasible: by s.10(3) it is subject to s.38(3), under which an order may apply the use-of-premises offence to betting of that kind.
- Gambling Commission (2026). *Prediction Markets – Here’s What You Need to Know*. Blog post by Brad Enright, Director of Strategy. Birmingham: Gambling Commission, 4 February.
- ICE (2025). *ICE Announces Strategic Investment in Polymarket*. Atlanta: Intercontinental Exchange, 7 October. Up to \$2bn, at a valuation of approximately \$8bn pre-investment.
- Johnson, K. (2025). *Farewell Address*. Washington, DC: Commodity Futures Trading Commission, 3 September. URL: <https://www.cftc.gov/PressRoom/SpeechesTestimony/opajohnson25> (visited on 3 September 2026).
- Kalshi (n.d.). *Member Agreement*. Section VI, “Representations and Warranties”, under which a member represents that they are not domiciled, organised or located in any listed prohibited jurisdiction; the United Kingdom is listed (visited on 1 September 2026).
- Kalshi (2026). *Fee Schedule*. kalshi.com. Taker fee: seven per cent of the number of contracts multiplied by the price and by one minus the price, rounded up on the order, with a default multiplier of one; no series charges a taker more, and a few series charge nothing. A formula of the same shape, without the multiplier term, appears in the rule certification Kalshi filed with the Commission in September 2022 (visited on 1 September 2026).
- KalshiEX v. Assad* (2026). No. 25-7516 (9th Cir., 28 August). Holding that sports event contracts are likely not swaps under the Commodity Exchange Act, and expressly disagreeing with *KalshiEX v. Flaherty*. Kalshi’s appeal was heard alongside those of Crypto.com and Robinhood; the court remanded Nevada’s challenge to election contracts.
- KalshiEX v. CFTC* (2024). No. 1:23-cv-03257-JMC, Document 51, US District Court for the District of Columbia (Cobb, J.), 12 September. Held that “gaming” means playing a game for stakes and that the public-interest review was not triggered; the Commission dropped its appeal on 5 May 2025.
- KalshiEX v. Flaherty* (2026). No. 25-1922, 172 F.4th 220 (3rd Cir., 6 April). Affirming a preliminary injunction by two votes to one on a likelihood-of-success standard, on both field and conflict preemption grounds. Not a decision on the merits. The Ninth Circuit heard the same question on 16 April 2026 and decided it on 28 August, as

- above; appeals remain pending in the Fourth Circuit, argued 7 May (*KalshiEX v. Martin*, No. 25-1892), and the Sixth, argued 30 July (*KalshiEX v. Schuler*, No. 26-3196); Kalshi appealed *Williams* to the Second Circuit on 8 July (No. 26-1835). *KalshiEX v. Williams* (2026). No. 1:25-cv-08846 (AT), US District Court for the Southern District of New York, 7 July, ECF No. 106. URL: <https://www.courtlistener.com/docket/71766515/kalshiex-llc-v-williams/> (visited on 1 September 2026).
- Murphy, A. H. (1973). “A New Vector Partition of the Probability Score”. In: *Journal of Applied Meteorology* 12.4, pp. 595–600.
- Park, D. (2026). “Kalshi and Polymarket’s Combined Volume Surges 75% to \$45 Billion in June Amid World Cup Fever”. In: *The Block*, 1 July. Kalshi \$31.5bn; Polymarket \$13.3bn across its two products.
- Polymarket (n.d.). *Geographic Restrictions*. Help centre. The United Kingdom is listed among the blocked jurisdictions; Polymarket’s developer documentation classes it as close-only, permitting existing positions to be closed but not new ones opened (visited on 1 September 2026).
- Polymarket (2026). *Fees*. Developer documentation, docs.polymarket.com. Taker fees first appear in the changelog on 5 January 2026, on fifteen-minute crypto contracts; they extend to college basketball and Serie A from 18 February, to all crypto from 6 March, and to a full category schedule on 30 March, which still leaves geopolitical markets fee-free. No fee entry appears earlier, and the log runs back to 15 May 2025. Current taker rates run from zero to 0.07 by category; makers are never charged (visited on 2 September 2026).
- Twycross, F., Baroness Twycross (2026). *Gambling: Regulation*. Written answer HL14169. London: Department for Culture, Media and Sport, 16 February.

The US-EU Productivity Paradox: The Debate Around Measurements

Sohan Sahu

Introduction

Over the past few months, an interesting debate has emerged over how to measure relative productivity growth between the USA and Europe, resulting in what has been described as a “US-EU productivity paradox.” This has drawn in many eminent and well-respected economists including Luis Garicano, Antonin Bergeaud, Brad DeLong, Robert Inklaar and Nobel laureates Paul Krugman and Philippe Aghion.

The debate centres around a divergence observed by Seth Ackerman in his Substack blog earlier this year, which showed conventional constant PPP measures of European productivity, as used in the 2024 Draghi report, declining relative to America, but current PPP measurements showing European productivity remaining relatively constant to America. This divergence is seen in *Figure 1* below:

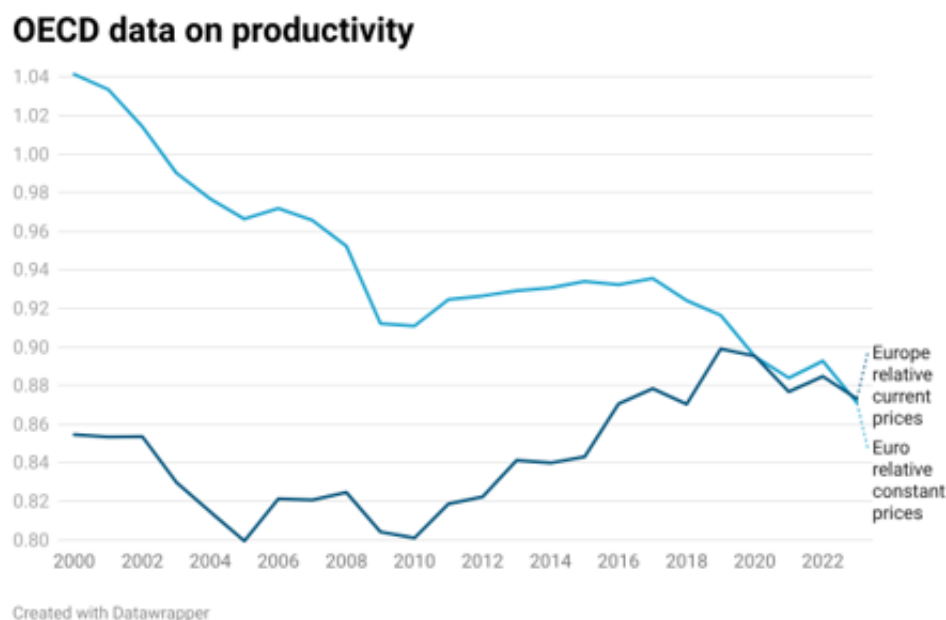


Figure 1: *European GDP per hour worked relative to US, between 2000 and 2023, PPP converted for constant prices (Base year 2020) and current prices (Krugman 2026c)*

As seen in *Figure 1*, the two measures give very different pictures of European productivity relative to the USA – at constant prices (light blue) European productivity relative

to the USA has declined between 2000–2023, but at current prices (dark blue) European productivity relative to the USA has been largely unchanged, which would suggest American productivity growth has not substantially diverged from Europe’s.

This has become the crux of the debate: Krugman, DeLong and more argue that the constant chain-linked PPP measure underestimates Europe’s relative economic welfare, and that interpretations of this give an incorrect assessment of Europe’s relative levels to the US. They argue this divergence is due to the way GDP is calculated, and that the different sectoral compositions of the US and European GDP figures mean constant PPP does not account for Europe benefiting from cheaper American technology. Aghion, Bergeaud, Garicano and many more argue the converse, claiming that current PPPs are an inaccurate measure of Europe’s relative productivity, because the current PPP transformations have a residual, which obscures a divergence in real growth in the USA relative to Europe. They argue this residual is due to various reasons, and as a result, current PPP reprices output every year, changing ‘the yardstick’ and, hence, conflates real volume changes with price changes. To put it more simply, as Inklaar does, “the core issue is divergence between constant and current PPPs is due to European price levels falling relative to US by more than the inflation differential would imply,” (Inklaar 2026).

The general agreement is that this divergence has to do with the rapid productivity gains and quality improvements in the American technology sector, echoing the famous remarks of Robert Solow (1987) that “You can see the computer age everywhere but in the productivity statistics,” reflecting how difficult it remains to measure productivity in this sector, nearly 40 years on. This is seen in the findings of a letter by the Federal Reserve Bank of Chicago, who found the IT sector to only be $\sim 8\text{--}9\%$ of US value added while driving roughly 45% of US total factor productivity growth since the late 1980s (David et al. 2025). This productivity growth, though, has resulted in a decline in the price of a unit of value added in the IT sector: it has fallen by more than 60% over the 1988–2023 period (David et al. 2025). This is seen in *Figures 2 and 3* below and reflects why it has been so difficult for statisticians to accurately measure productivity growth, hence the ongoing debate.

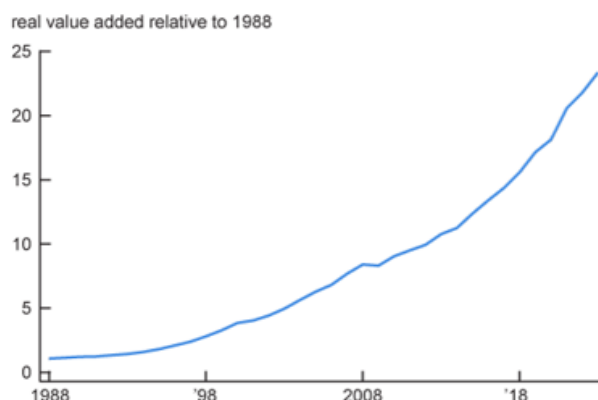


Figure 2: *Information technology (IT) sector cumulative real value-added, 1988–2023* (David et al. 2025)

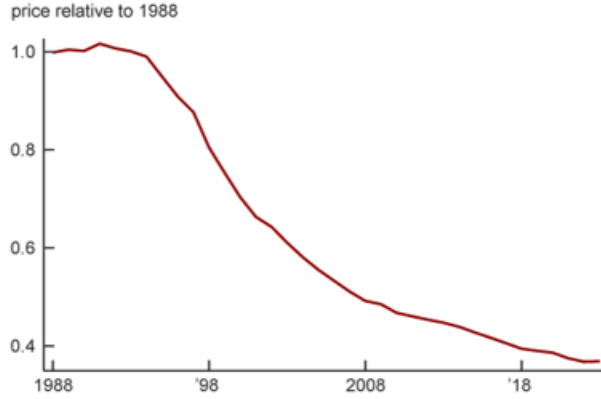


Figure 3: *Information technology (IT) sector price of a unit of value added, 1988–2023* (David et al. 2025)

The intention of this article will not be to provide a judgement on this debate, which is clearly dividing many great economists, but instead to offer a general summary of the detailed arguments made. The hope is that this will offer a stepping stone for readers to conduct their own research, and come to their own, informed conclusions, while highlighting the complexity and nuances in calculating these statistics.

What is agreed upon

Before starting a detailed breakdown of the main arguments, it is worth clearly highlighting the points the two main sides in this debate do not dispute, to avoid any confusion or misrepresentation of arguments later on. The most significant one is the agreement that US productivity growth has in fact outpaced Europe, with Krugman (one of the main proponents of the current PPP measure) himself declaring, “I am not arguing that European productivity is mismeasured, and never said that,” (Krugman 2026b). This is evident across data sources, as Aghion et al. show in their response to Krugman in *Table 1*:

Source	France (%/yr)	United States (%/yr)
National accounts (INSEE and BEA)	+0.85	+1.93
EU KLEMS	+0.96	+2.13
OECD	+0.94	+1.74
Bergeaud, Cette, and Lecat	+0.80	+1.71
Gethin and Saez	+1.19	+1.61
Andreescu, Loubes, Piketty, and Robilliard	+0.91	+1.79
Conference Board	+0.79	+1.79
World Bank current PPP / PWT hours	+3.75	+3.89

Table 1: *Average annual growth in labour productivity in the USA and France from 1995 onwards. The first seven rows show real GDP per hour worked, while the last row shows output per hour when GDP is converted using current PPPs* (Aghion et al. 2026)

The debate centres around the final row, which Aghion et al. argue is not an accurate reflection of relative productivity growth between France and the USA, while Krugman and more argue it is, because France also benefits from cheaper American technology, even if France’s productive capacity has grown relatively less. What is not disputed are the other sources, which consistently show US productivity growth year-on-year outpacing France, by about one percentage point. As explained in the introduction, both sides acknowledge this is due to the productivity growth in the American technology sector, and since technology is a tradeable good, Europeans can buy this technology at similar prices to Americans. The dispute then also centres around how the productivity gains are distributed, and whether this is enough to offset the underlying growth differential in economic welfare comparisons between Europe and the USA.

There is also a recognition that both measures (constant and current PPP conversions of GDP) are imperfect, because of the difficulties in calculating statistics for temporal and spatial comparisons. James K. Galbraith references The Economist’s famous Big Mac index, as an example, to illustrate the difficulties of creating an accurate PPP comparison: The Big Mac index suffers from different expectations and customer profiles at McDonald’s, as well as different taxes, e.g. VAT and social spending, across different countries (Galbraith 2026). However, as Dean Baker shows, national deflators also struggle with the difficult problem of hedonic adjustments (Baker 2026). He illustrates this with the example of mobile-phone price indexes, which fell 18% in Australia but 52% in the US and Canada and 90%+ in the UK and New Zealand over 2008–2018, for what is nominally the same technology (Byrne 2019). While there is an acceptance that PPPs are better for spatial comparisons and national deflators for temporal comparisons, asking either to do both simultaneously is often challenging and leads to a lot of statistical difficulties.

The arguments in favour of current PPP measurements

As has been mentioned earlier, the key argument of this side of the debate, as said by Paul Krugman, is “whether the standard comparison of European and US productivity growth rates is a good indicator of what is actually going on in the two economies over time,” (Krugman 2026b). Their argument hinges on the idea that constant PPP is not a good measure of relative economic welfare, for three reasons. 1.) It does not capture how productivity gains are distributed and diffused through lower prices, which benefits Europeans since tech is tradeable; 2.) Constant PPP comparisons require extrapolating from the base year using national deflators/growth rate statistics, but these are plagued by inconsistencies in hedonic adjustments across countries; 3.) Constant-price PPPs extrapolated through deflators are not very accurate, and produce unusual level claims, depending on the choice of base year.

For the first aspect, DeLong explains how the “product wage” (nominal wage $\div$ price of what a country produces) has risen enormously for US tech producers, as tech prices collapse from productivity gains, but the “user-consumption wage” (nominal wage $\div$ price of a broad consumption basket) has not, because tech is a small share of what anyone actually consumes. Hence, European “users” of high-tech have seen their real earnings rise about as fast as American users, even though Europe does not produce these technologies (DeLong 2026). Or to put it in other words, if the assumption that American technology markets are competitive holds true, then the productivity gains

from the technology sector are not captured as profits by producers, but instead passed through to consumers as lower prices. Since these are tradeable goods, Europeans also benefit from these lower tech prices, meaning Europe’s real purchasing power has also increased.

Krugman elaborates on this further, with a basic stylised model (Krugman 2026d):

- Assume there are two goods, tech (T) and nontech (N), both of which are tradable and have no additional costs associated with trade.
- Also assume that consumers in both countries share Cobb-Douglas preferences, spending a constant share $\tau < 1/2$ of their income on T.
- Then, assume labour productivity in the production of N is the same in both countries, and productivity growth in the sector is 0 in both countries.
- Assume the US has a comparative advantage in the production of T, such that all T is produced in the US and that productivity growth in this sector is at a rate ρ .
- Then this model makes a couple of predictions: Wages in the nontech sector are the same in the EU and US, because N is a tradable good; Tech will attract a share τ of world spending and hence, will account for a share τ of world GDP, and 2τ of US GDP because $\text{World GDP} = \text{US GDP} + \text{EU GDP} = 2 \text{ US GDP} = 2 \text{ EU GDP}$.
- Technological progress in T implies rising US real GDP, measured as growth in “chained” constant prices, at a rate of $2\rho\tau$. Growth in EU real GDP is zero. Yet relative GDP at current prices remains 1. This is because real GDP growth from tech volume growth is at a rate ρ , but consumers only spend a fixed share τ on tech, so absorbing a $1 + \rho$ rise in tech volume at unchanged nominal spending requires tech prices to fall by ρ , leaving total nominal (current-price) GDP unchanged.
- Additionally, because nominal wages (equal in both countries) are pinned down by zero-productivity growth in the non-tech sector, the cost-of-living falls by $\rho\tau$ (spending share on tech $\times$ price fall). As a result, real wages will rise in both countries by $\rho\tau$.

The key consideration in this chain of reasoning is the assumption that the American technology sector is competitive, and this not only has implications for Europe, but also for the distribution of productivity gains between producers and consumers in the USA. Krugman highlights two sources to argue that these productivity gains have been diffused through lower prices, and not higher wages or profits for producers of American technology. Firstly, he cites a comparison of estimates for mean household income by the LIS (a cross-national data center in Luxembourg) showing that between 2000 and 2021, the data showed nominal income rising 3.1 percent annually in the Netherlands compared to 3.3 percent in the US (Krugman 2026a). The implication of this is that there has not been an observable divergence in mean household income that would be suggested if the productivity gains in the USA had been translated into higher wages for workers in the technology sector. Secondly, in an internal comparison within the US, Krugman highlights how “according to the Bureau of Labor Statistics, output per worker hour in computer manufacturing has risen by a factor of more than 14 since 1988, or about 10 percent per year. As best as BLS can measure it, output per hour in hospitals barely changed over the same period. But we didn’t see a correspondingly huge rise in the wages of computer workers relative to those of doctors and nurses,” which once again

implies that workers in the technology sector have not seen their wages rise, as would be expected if the productivity gains from the sector were captured by producers (Krugman 2026a). There is less debate, though, around the fall in prices of technology in America – an example of this can be seen earlier in *Figure 3*.

Next, as Ackerman noted in his original post, one potential channel to explain this divergence is the difficulty national deflators face with hedonic adjustments. Krugman (2025) illustrates this problem with a simple example highlighting how cutting-edge CPUs today process 1000s of times more instructions than in the 1990s, but productivity in the computer sector has not risen by 1000x+ because processing more data does not make it better, causing difficulties for statisticians who need to account for quality changes when tracking CPU prices over time. This does cause issues with real-world stats, as Ackerman (2026) cites a paper showing that while television-set prices declined by 85% in the US (BLS) price index over 2005–2015, they increased by 8% in Japan’s official statistics (Shiratsuka 2023). This is important because the constant PPPs extrapolate from a base year using national deflator statistics, so the visible inconsistencies in the hedonic adjustments could cause divergences in spatial comparisons, giving misleading pictures of relative productivity. This problem does not affect PPPs to the same extent, because they only compare standardised products across countries, and under the current-price comparisons, products are repriced every year, taking into account the quality adjustments.

There is also further evidence that the problems with hedonic adjustments may stretch beyond just a few products in the IT and technology sectors. In a recently published working paper, the authors compare consumer to producer price indices, and using an input output framework, they find “that TFP growth is understated by 1.4 percentage points in durable manufacturing and 0.4 percentage points in nondurable manufacturing, and is slightly overstated in nonmanufacturing industries,” (Atalay et al. 2026). As a result, they conclude that the observed stagnation in TFP growth in manufacturing industries in the USA, post-2000s, is likely a statistical artefact, because conventional measures have failed to fully capture quality improvements. While this study only covers the USA, it reflects a more general problem faced by hedonic adjustments. If it is consistent across borders, then since manufacturing is a larger share of Europe’s GDP compared to the US, higher productivity growth in manufacturing sectors would result in a smaller differential to the US for Europe.



Figure 4: *OECD estimates of ratio of GDP per worker-hour in the US and Netherlands, adjusted for year-on-year current price PPP and base year 2020 PPP prices adjusted for national inflation rates (Krugman 2026b)*

For many, though, the most interesting argument on this side of the debate highlights the unusual results the constant PPP measure causes. As Krugman explains using *Figure 4*, while under current-price PPP the Netherlands is slightly more productive (probably because of capital-intensive industries around the port of Rotterdam), using chain-linked prices shows the Netherlands was 25% more productive than the USA in 2000. Krugman argues this is *reductio ad absurdum*: It would go against conventional understanding, in stating that the Netherlands was significantly more productive than the USA in 2000 (Krugman 2026b). He goes on to argue this is because using 2020 prices as a base year massively undervalues actual US output in 2000, because technology prices are far cheaper, which creates the big gap observed. In a postscript, though, Aghion, Bergeaud and Garicano indicate they are willing to accept this level claim. However, as observers note, OECD data produces even more unusual statistical results: If constant prices are used with 2020 as the base year, then this leads to Spain being more productive than the USA in 1995, and Italy being 20% more productive than the USA in 1995, which means the US would have been one of the poorest and least productive countries in Western Europe at the time (OECD 2026). While this may be due to specific circumstances which distorted Spanish and Italian productivity figures, the general level results run against the prevailing economic consensus.

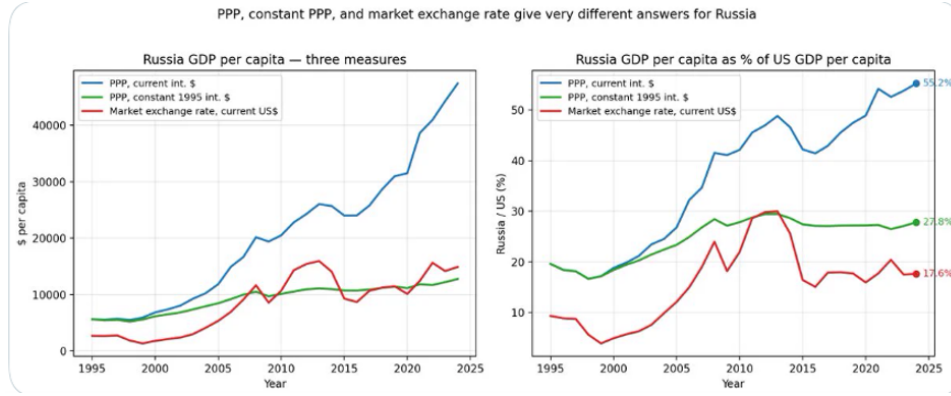


Figure 5: *PPP, Constant PPP and market exchange rate for Russian GDP per capita individually and as % of US GDP per capita between 1995–2024, sourced from Penn World Tables (@HannoLustig 2026)*

As Hanno Lustig notes on X though, these statistical oddities also appear in the current PPP measures, offering a similar *reductio ad absurdum*. As seen in *Figure 5* above, using current PPP shows a “Russian growth miracle,” suggesting it has nearly quintupled GDP per capita in the past 30 years (@HannoLustig 2026), which again does not align with the current economic consensus, and is instead likely a statistical artefact caused by the relatively low prices of certain products in Russia, combined with the Russian wartime economy.

The arguments in favour of constant PPP

For the most part, the leading rebuke against Krugman and more has been led by Bergeaud and Garicano, who, like others, have raised their biggest objection as the repricing of output every year by current PPPs, which gives a misleading image of Europe’s relative productivity growth, by obscuring differences in relative productivity gains in the US. As they note, “when the PPP path and the national deflator path diverge, current PPP growth mixes changes in quantities with changes in the prices used to value those quantities,” (Aghion et al. 2026). To put it more simply, current price comparisons miss productivity gains in sectors where prices fall, as they show with an anecdote where if the USA produces 2x more software, but price halves, then at current prices the value of software output is unchanged, despite the real growth where volume has doubled.

$$G_{AB,t} = \log\left(\frac{\text{PPP}_{AB,t}}{\text{PPP}_{AB,0}}\right) - \log\left(\frac{D_{A,t0}}{D_{B,t0}}\right).$$

Equation 1: *Identity detailing the deflator-PPP gap (Bergeaud 2026)*

In an accompanying technical note, Bergeaud explains this residual in more detail with the identity seen above in *Equation 1*: the first term reflects the residual between the PPP measurement and deflator; the second term consists of the spatial price comparison, which reflects the difference in the cost of living between countries (Country B has PPP standardised at 1, hence it is only PPPs of A at time t and t=0); the third term contains the domestic deflators, reflecting the difference in inflation rates at time t. The exact

derivation of *Equation 1* can be found in the technical note, but this identity formally states that if the PPP ratio declines relatively more than the deflator ratio, then the residual becomes negative. Bergeaud explains that this residual could cause the current PPP comparison to make country A’s relative growth look stronger than the deflator ratio indicates (Bergeaud 2026).

$$d \log PPP_{AB,t}^T = (d \log P_{At} - d \log P_{Bt}) - \frac{1}{2} \sum_k (s_{kAt} - s_{kBt}) (d \log p_{kAt} + d \log p_{kBt}).$$

Equation 2: Gap between country A and B PPP, derived from Equation 1 and a bilateral Tornqvist spatial comparison of countries A and B. See the technical note for full explanation (Bergeaud 2026)

Between the technical note and the response to Krugman, there are two key reasons identified that cause this residual and hence suggest current PPPs are a less accurate measure. The first reason, referenced earlier, is to do with the different weightings of sectors for PPP and national deflator calculations. This is demonstrated in Bergeaud’s technical note, where starting with the original identity stated in *Equation 1*, he shows it is equivalent to compare current PPP with a deflator ratio as to compare current PPP with the PPP anchored to a base year and extrapolated using that deflator ratio (which is the main comparison seen throughout). Then, using the work of Deaton and Aten (2017), he constructs a bilateral Tornqvist spatial comparison of country A with B, effectively standardising a common PPP unit as the average price level of country A and country B, weighted equally. The resulting identity is seen below in *Equation 2*, but importantly, it has an additional term that reflects that changes in current PPP can also be due to a combination of the expenditure share of any sector in country A and B differing, when there is a non-identical price movement, causing the current PPP metric to change. More intuitively, if a country has a lot of expenditure in a sector whose prices are rapidly falling, then its deflator will fall significantly, but because PPP is weighted for both countries, the PPP does not fall as rapidly, and so will not register the same scale of change as the national deflator (Bergeaud 2026).

This is also exaggerated by the fact that international PPP comparisons use multilateral spatial formulas, to ensure transitivity between countries, because PPP calculations using country-specific weights are usually inconsistent with transitivity (Hill 2004). As a result, the weights used by the ICP (International Comparison Program – which is the main database for PPP calculations) depend on the entire system, and not one country only (Deaton and Aten 2017). This is a specific reflection of the more general difficulties PPP calculations face, because the statistical agencies often do not have data on the weighting of the expenditure on each individual item, so they are often forced to estimate the weights (unlike national statistical agencies), but this can make items have unrepresentative weightings in regard to consumption patterns in different countries e.g. baguettes in France would be a common daily purchase, but in the USA would not, and so attempting to standardise a weight for baguettes can create unrepresentative weightings within each category.

The second issue, which Bergeaud flags as the bigger contributor to the residual, is product sampling issues faced by PPP calculations. This is because, to actually calculate

an accurate PPP rate, indices are forced to adhere to strict, standardised products with easy cross-country comparisons, but this creates unrepresentative baskets of domestic consumption. As Inklaar (2026) shows, the difference is due to the products chosen, because these are often US technology goods, where prices are collapsing from productivity gains. By contrast, national deflators used in the constant PPP calculations consider how a country’s relevant output or expenditure components change over time. The example of computers is once again used here by Bergeaud (2026), since national deflators would track computer prices through quality changes, such as improvements in memory and processing speed, while PPP metrics price a computer based on a specified model between two countries that can be easily compared at the same date. This difference is even clearer with healthcare, because while national deflators can track the price of the same healthcare bundle in a country over time, PPPs must account for the different institutional settings including insurance, payment rules, providers and much more. The result is that the products chosen for PPP are matched between different countries, while products for national price indices are different as they track price changes over time, meaning the ICP price lists (used by PPP) and CPI price lists (used by national deflators) contain explicitly different products (Deaton and Aten 2017).

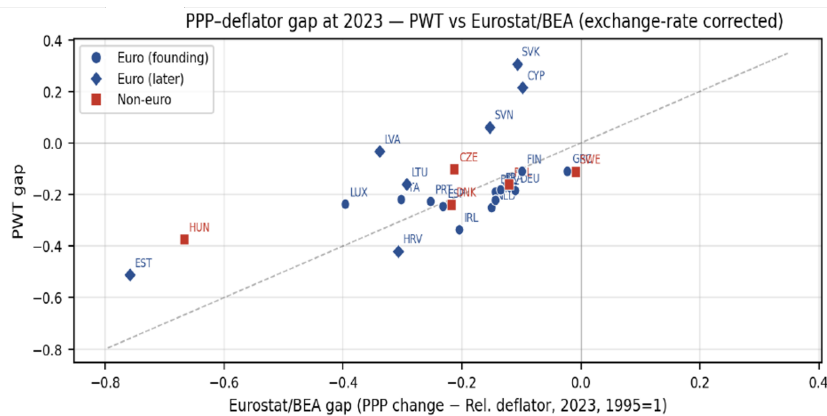
It is also worth briefly considering some important rebuttals made to Krugman’s stylised model, which argue the assumptions made by Krugman are in fact unrealistic. For one, economies have non-tradable sectors like healthcare and hospitality, meaning wages do not equalise between economies from trade, since higher wages in the US technology sector spill over into these non-tradable sectors, whose wages are not tied to any European counterparts. Secondly, it is argued that the technology sector is not in fact a competitively priced market following marginal-cost pricing, as they reference “40% margins” for Apple or “reported 70% margins on inference” for Anthropic, as evidence that most productivity gains stay as profits for American producers, not as lower prices (Garicano and Garicano 2026).

Recent research also provides evidence to refute the claim that any growth differential between the US and Europe is due to Americans increasing their working hours relative to Europeans. In fact, the paper by Birinci et al. (2026) shows the converse, with the gap between working hours of Americans and non-Americans falling by roughly half between the 1990s and the 2010s. The authors’ findings are that this is to do with the rapid expansion of Medicaid benefits in the USA, which caused working hours to decline in the US, while Europeans and others did not see the same relative change in benefit generosity, causing working hours to remain largely unchanged or even increase a little (Birinci et al. 2026).

What the data says in this debate

With the main arguments outlined above, Robert Inklaar, one of the leading economists in the world on international growth and productivity comparisons, authored a detailed data-driven breakdown of this debate. Using the same identity as Bergeaud in *Equation 1*, Inklaar (2026) explains “the change in a country’s relative price level measured at current PPPs equals its real-growth differential with the US minus the gap; equivalently, the gap is the relative real-growth differential minus the change in the current PPP relative level.” From his data analysis, he finds three key takeaways: Firstly, the gap

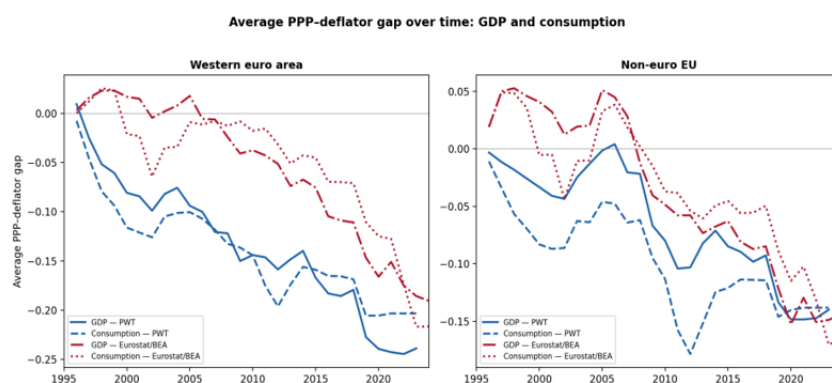
between the deflator and PPP measurements is consistent across the two data sources used (Eurostat and PWT). Secondly, this gap is consistent across different variables, namely GDP and consumption, as well as the relevant adjustments to them. Finally, disaggregating the PPP-deflator gap by category shows that while prices in PPP measures fall faster for tradeable goods than for deflators, prices for deflators fall faster than PPP for non-tradable goods (Inklaar 2026).



GDP, 1995 = 1. Circles = founding euro members; diamonds = later joiners; squares = non-euro. Eurostat/BEA deflators converted to national currency.

Figure 6: *PPP-deflator gap in 2023 for selected countries in Europe, using data from Penn World Tables, Eurostat and Bureau of Economic Analysis (Inklaar 2026)*

Figure 6, above, outlines that the PPP-deflator gap is consistent across all European countries, and consistent across the two data sources. Notably, the vast majority of European countries sit in the third quadrant, indicating both the Eurostat/BEA data and Penn World Tables show a negative gap between PPP and national deflators – the few that are not in this quadrant are economies where the PWT gap is positive, and are all small economies where PPP measurements are inherently noisier (Inklaar 2026). What this highlights is that the PPP-deflator gap is due to PPP price indices falling more than national deflator indices, and that this is a general phenomenon, not a statistical anomaly between comparisons of a few countries in a single data source.



Unweighted average across countries. Western euro area: BEL, DEU, FRA, ITA, ESP, NLD, AUT, FIN, IRL, PRT, GRC. Non-euro EU: DNK, SWE, HUN, POL, CZE. Eurostat/BEA deflators converted to national currency for non-euro countries.

Figure 7: *The average PPP-deflator gap between 1995–2023 for two groups of European countries: Countries in western Europe who use the Euro, and countries in the EU, who do not use the Euro, based on data from Penn World Tables and Eurostat (Inklaar 2026)*

Figure 7 shows Inklaar’s second observation: That the PPP deflator gap is also generally negative over time for consumption, as well as for GDP. As Inklaar (2026) notes, *Figure 7* illustrates that the “GDP PPP-deflator and consumption PPP-deflator gaps track each other pretty well, especially in Western Europe, starting at around 0 in 1995 and ending up between -0.15 and -0.25 by 2020s.” This is important in this debate, because it implies that the driver of the PPP-deflator divergence operates on consumption, just as much, as it does on GDP (Inklaar 2026). This observation, hence, negates one of the key hypotheses of Krugman in favour of current PPP measures, because it suggested that the GDP deflator would diverge from GDP adjusted for PPP, and hence mean European consumption prices would begin to converge with American consumption prices. This would have been due to Europeans consuming cheaper technology, while producer prices would not converge, but instead, the consumption PPP deflator gap is not zero, and the countries with relatively larger GDP-deflator gaps also have relatively larger consumption PPP-deflator gaps.

PPP change and relative HICP change by COICOP category, 2002–2021 (Western euro area average, 2002 = 1). Gap = PPP change – Rel. HICP.

COICOP category	Rel. HICP	PPP change	Gap
01 Food & beverages	0.940	0.869	−0.071
03 Clothing & footwear	1.087	0.845	−0.242
04 Housing & utilities	0.915	0.861	−0.054
05 Furnishings	1.204	0.933	−0.271
06 Health	0.752	1.281	+0.529
07 Transport	1.006	0.746	−0.261
08 Communication	0.727	0.501	−0.225
09 Recreation & culture	1.031	0.723	−0.308
10 Education	0.662	1.094	+0.432
12 Miscellaneous	0.863	0.818	−0.045

Table 2: PPP by consumption category between 2002–2021, averaged for the western European countries using the Euro (Inklaar 2026)

The most interesting finding of Inklaar, though, is shown above in *Table 2*, where he uses bilateral Fisher PPPs and the Harmonised Index of Consumer Prices (HICP) to disaggregate the gap by category (sector). As seen in *Table 2*, “for categories dominated by internationally traded goods, PPPs have declined substantially more than relative HICP inflation would predict,” as seen with the gap in recreation and culture, where PPP fell to 0.72 times its 2002 levels, while the HICP measure was mostly unchanged at 1.031. There were also similar relative declines for PPP compared to HICP, for furnishings and transport, while there was a smaller negative gap for communication and clothing (Inklaar 2026). Notably, these are sectors dominated by rapid technological change and are often highly tradeable on world markets, suggesting PPP measures have found it easier to account for technological changes since they only need to compare largely standardised goods across borders, while national price deflators need to track price changes and quality improvements. This is reflected in *Figure 8* below, which shows that some of the sectors with the largest contribution to US TFP growth are technology and retail trade – sectors where PPPs have declined relatively more than the HICP prices.

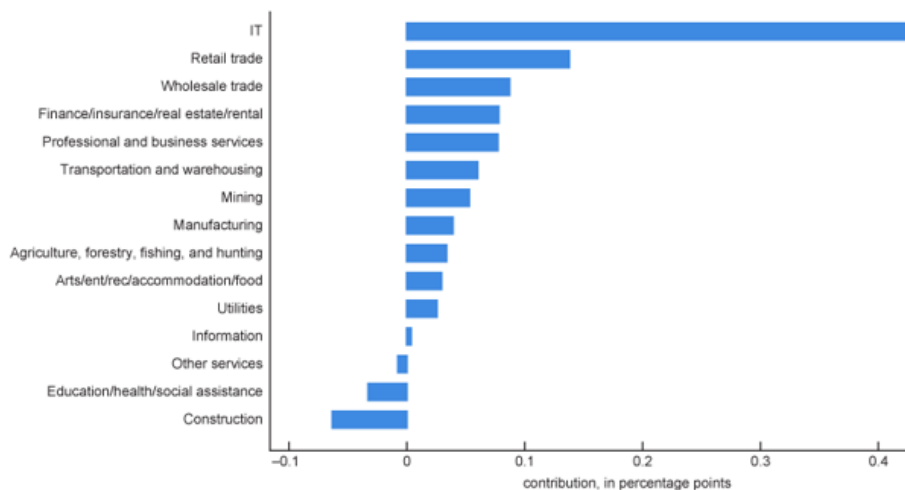


Figure 8: *Sector contribution to aggregate TFP growth in the USA between 1988–2023* (David et al. 2025)

By contrast, in service-dominated categories, such as healthcare or education, the gap between PPP and HICP is reversed in Table 2. For example, in healthcare PPP prices rose to 1.28 times their 2002 level, while HICP fell to just 0.75 times its 2002 level. There is a similar gap in education too, and this is likely because output in these sectors is heterogeneous and varies a lot in quality, making it difficult to measure. While PPPs have to compare baskets that are incomparable across different institutional settings in different countries, national deflators only need to track a single institutional framework over time and so, while imperfect, are probably more accurate, suggesting “that cross-country PPP comparisons are less reliable than time-series deflators for these hard-to-measure service categories,” (Inklaar 2026). What this implies is that the main reason for this PPP-deflator gap is the product sampling issue explained earlier, but as seen here, it is direction-neutral, since it varies between sectors, and is not a weighting issue since “a synthetic PPP built from a fixed representative basket and rolled forward on each country’s own domestic price changes reproduces that country’s national deflator,” (Inklaar 2026).

As Inklaar (2026) explains, even though these two effects (negative in tradeable sectors, and positive in non-tradeable sectors) partly offset in the aggregate, the net PPP-deflator gap for Europe remains substantially negative as seen in *Figures 6 and 7*. This leads him to conclude that the growth differential between the US and Europe is the most reliable and model-independent part of the data “because the measurement error around any single PPP is larger than that around growth rate,” with the question of the relative gap of the levels (if there is any at all) left unresolved, since it depends entirely on an effectively arbitrary choice of base year. In fact, Inklaar suggests that opting for a base year in the past (e.g. 2000) and then extrapolating (as some constant PPP comparisons do) is an unproven burden so far: It either requires the PPP conversions in the chosen year to be superior to those in the present, or for the structure of the basket to be stable enough for it not to matter, but this is not the case for the categories driving the PPP-deflator gap, such as the communication, recreation and technology sectors. A similar conclusion was reached by former IMF Chief Economist Olivier Blanchard, who agrees that constant-price chain-linked PPPs should be favoured, but that there is no concrete

conclusion on the level results (@ojblanchard1 2026).

What are the alternative interpretations of this debate?

While the arguments and data outlined earlier cover the main parts of this debate, several well-respected observers have offered other perspectives on this debate. James K. Galbraith questions how well GDP reflects living standards at all. He argues that the average American is “over sixty percent richer, in real terms, in 2026 than in 2000”, while real median weekly earnings, according to the St Louis Federal Reserve, have risen by only 10.5% over the same period (Galbraith 2026). The St Louis Fed’s own series for real GDP per capita shows a smaller rise, of roughly 44% between 2000 and 2026 (U.S. Bureau of Economic Analysis 2026), but that still leaves a wide gap with median earnings. Galbraith sees this as a broader disconnect between GDP figures and real median weekly earnings, to the extent that he claims GDP figures must be very distorted by the skewed income distribution in the USA.

Rob Gilles (2026) goes further, arguing in his blog that the debate is flawed because it is conducted in the wrong units: It accepts GDP as the right measure for productive capacity and relative economic welfare. On the production side, because GDP weights output by market price, it only becomes a valid welfare measure under perfect competition, but Gilles argues that the reported margins in the technology sector (approximately 40% for Apple, and 70% on inference for Anthropic) and the fact that countries that host IP registrations like Ireland can inflate GDP, with accrued rents, mean GDP is not an accurate assessment of real output (Gilles 2026). As Gilles notes, PPP conversions are also flawed for reasons explained earlier, and so he proposes future measurements based on a wholly theoretical “standard commodity” — a composite good built from actual physical inputs/outputs — that would strip out price-cost margins and IP-domicile effects to isolate genuine productive capacity (Gilles 2026). However, at this moment, this measure does not exist, and given the difficulties seen in calculating PPPs, it would likely require an immense amount of statistical work to make it widespread, accurate and reliable.

Finally, there is the idea that this debate has missed the demand side of the picture, which is laid out by Dimitri Zenghelis in an FT Opinion Letter. He argues that “as people grow richer, they spend proportionally more on things like recreation, fashion, financial and professional services, health and social care — sectors in which Europe excels — and relatively less on manufactured goods. These services see little ‘real’ productivity growth, but their prices rise anyway in step with growing demand — a Baumol-style effect. The underlying shift in preferences allows them to pay wages competitive with high-productivity manufacturing, without needing manufacturing’s productivity gains to justify them,” (Zenghelis 2026). The Baumol effect, simply put, is the rise in wages in sectors with low-productivity growth, so they can compete with rising wages in high-productivity sectors, resulting in these sectors becoming more expensive as input costs rise faster than productivity — Zenghelis flips this narrative to argue that wage rises are due to higher demand, stemming from productivity improvements’ effects on the demand side, where they drive up wages.

Conclusion

At the beginning of this article, it was set out that this piece should not provide a judgement on this highly complex and technical debate. Instead, it has hopefully highlighted the difficulties of index calculations and translating them for use in both the spatial and temporal dimensions. This is a very relevant and crucial question though, because as seen in the introduction, the divergence can lead to very different framings of European economic performance relative to the USA – under constant PPP measures Europe has seen relative productivity to the USA decline, but under current PPP measures relative productivity has remained largely unchanged. This can influence policymaking in the real world: The Draghi report (2024) to the European Commission used the constant PPP measure to present a narrative of relative underperformance by the European economy. Meanwhile, DeLong (2026) likened the current PPP measurements to the Lewis-Prebisch-Singer framework, which established that the benefits of increased productivity in primary products were mainly captured by consumers, while growth in manufacturing productivity was mainly retained by producers – effectively diagnosing that the statistics do not depend on technological stagnation, but the allocation of surplus gains from the American technology sector, and hence are not a sign of relatively lower productivity growth in Europe.

Unfortunately, though, there is also no clear right side to this debate. While PPP calculations face challenges in attempting to create standardised, internationally comparable, and representative baskets of goods, national deflators often have difficulties with complicated hedonic adjustments. Ideal measurements, like the “standard commodity,” are not present and transferable to the real world, with the result being statistical oddities, such as the “Russian growth miracle” or constant PPP level claims. Inklaar (2026) probably offers the best objective assessment of this debate with his data-driven breakdown suggesting the compositional argument used to explain the divergence does not align with the data. However, Inklaar admits that while the existence of a growth differential is largely uncontested, it is the size of this differential and the level results that are highly uncertain.

The reality is the debate will likely linger for a while, while deeper research is conducted into the technicalities of these statistics, in an attempt to explain the divergence. This debate, though, sits in the context of a much broader one about the comparison of the standard of living in the USA and Europe. As Grossman and Helpman (1991) showed with their “quality-ladders” trade model, it is possible for a country which innovates less to still gain from trading with a technologically dynamic partner: competition forces the innovator to pass productivity gains through as lower prices, benefiting both sides. As Krugman (1991) explained though, agglomeration effects — the clustering of suppliers, specialised labour, and forward/backward linkages around the innovating country – can entrench a country’s productive capacity and pull future rounds of innovation away from lagging countries, rather than letting them catch up. Hence, while American productive capacity is undoubtedly growing faster than Europe’s, the question of how much this matters for present and future economic welfare and living standards comparisons is a discussion that could fill endless pages and requires far more metrics than just GDP and productivity.

References

- @HannoLustig (2026). *Russian Growth Miracle?* X post. URL: <https://x.com/HannoLustig/status/2054434119275417875> (visited on 30 July 2026).
- @ojblanchard1 (2026). *[Untitled]*. X post. URL: <https://x.com/ojblanchard1/status/2061775161955692551> (visited on 17 September 2026).
- Ackerman, S. (2026). *Europe’s Productivity Keeps Outpacing the US*. URL: <https://sethackerman.substack.com/p/europes-productivity-keeps-outpacing#footnote-1> (visited on 29 July 2026).
- Aghion, P., A. Bergeaud, and L. Garicano (2026). *The Mismeasurement of European Productivity (Revised)*. URL: <https://www.siliconcontinent.com/p/the-mismeasurement-of-europes-productivity#footnote-1> (visited on 29 July 2026).
- Atalay, E. et al. (2026). *Why Is Manufacturing Productivity Growth So Low?* Tech. rep. Cambridge, MA: National Bureau of Economic Research.
- Baker, D. (2026). *The Grand Illusion: The US–Europe Growth Gap*. URL: <https://cepr.net/publications/the-grand-illusion-the-u-s-europe-growth-gap/> (visited on 29 July 2026).
- Bergeaud, A. (2026). *Purchasing Power Parities and National Deflators Across Space and Time*. Tech. rep. Paris: HEC Paris.
- Birinci, S., L. Karabarbounis, and K. See (2026). *Why Do Americans No Longer Work So Much More Than Non-Americans?* Tech. rep. Cambridge: National Bureau of Economic Research.
- Byrne, D. (2019). *The Mysterious Cross-Country Dispersion in Mobile Phone Price Trends*. Tech. rep. Washington, D.C.: Federal Reserve Board.
- David, J., G. de Souza, and A. Schommer (2025). “Concentrated Growth: The Role of the IT Sector”. In: *Chicago Fed Letter* 515, pp. 1–9.
- Deaton, A. and B. Aten (2017). “Trying to Understand the PPPs in ICP 2011: Why Are the Results So Different?” In: *American Economic Journal: Macroeconomics* 9.1, pp. 243–264.
- DeLong, B. (2026). *Paul Krugman, Cui Bono from the Coming of the Bio-Info Tech-Attention Economy, & Embarrassing Conceptual Errors by Economists Who Really Should Know Better: Chart of the Day*. URL: <https://braddelong.substack.com/p/paul-krugman-qui-bono-from-the-coming> (visited on 29 July 2026).
- Draghi, M. (2024). *A Competitiveness Strategy for Europe (Part A)*. Tech. rep. Brussels: European Commission.
- Galbraith, J. K. (2026). *On the US-EU Debate: A Response to Paul Krugman*. URL: https://www.ineteconomics.org/perspectives/blog/on-the-us-eu-debate-a-response-to-paul-krugman#_edn7 (visited on 29 July 2026).
- Garicano, P. and L. Garicano (2026). *European Stagnation Is Real*. URL: <https://www.siliconcontinent.com/p/european-stagnation-is-real> (visited on 30 July 2026).
- Gilles, R. (2026). *The Big GDP Debate: Wrong Units*. URL: <https://www.relationaleconomy.net/2026/05/the-big-gdp-debate-wrong-units/> (visited on 30 July 2026).
- Grossman, G. M. and E. Helpman (1991). “Quality Ladders in the Theory of Growth”.

- In: *The Review of Economic Studies* 58.1, pp. 43–61.
- Hill, R. J. (2004). “Constructing Price Indexes across Space and Time: The Case of the European Union”. In: *American Economic Review* 94.5, pp. 1379–1410.
- Inklaar, R. (2026). *Comparing PPP Changes with Relative Inflation: Evidence from PWT, Eurostat/BEA, and Official PPP Statistics*. Tech. rep. Groningen: University of Groningen.
- Krugman, P. (1991). “Increasing Returns and Economic Geography”. In: *Journal of Political Economy* 99.3, pp. 483–499.
- (2025). *Europe’s Tech Lag: Does It Matter?* URL: <https://paulkrugman.substack.com/p/europes-tech-lag-does-it-matter> (visited on 29 July 2026).
- (2026a). *Challenging the Narrative of European Decline: Revised, Free Repost*. URL: <https://paulkrugman.substack.com/p/challenging-the-narrative-of-european-478> (visited on 29 July 2026).
- (2026b). *Europe Versus America: A Response to the Critics*. URL: <https://paulkrugman.substack.com/p/europe-versus-america-a-response> (visited on 29 July 2026).
- (2026c). *Europe Versus America: A Wonkish Data Follow-up*. URL: <https://paulkrugman.substack.com/p/europe-versus-america-a-wonkish-data> (visited on 29 July 2026).
- (2026d). *Modeling the US-Europe Paradox (Very Wonkish)*. URL: <https://paulkrugman.substack.com/p/modeling-the-us-europe-paradox-very> (visited on 29 July 2026).
- OECD (2026). *Productivity Levels*. URL: https://data-explorer.oecd.org/vis?lc=en&tm=productivity&pg=0&snb=585&vw=tb&df%5Bds%5D=dsDisseminateFinalDMZ&df%5Bid%5D=DSD_PDB%40DF_PDB_LV&df%5Bdg%5D=OECD.SDD.TPS&df%5Bvs%5D=1.0&dq=ESP%2BAUT%2BFIN%2BUS%2BITA.A.GDPHRS..USD_PPP_H.Q...&to%5BTIME_PERIOD%5D=false&pd=1995, (visited on 29 July 2026).
- Shiratsuka, S. (2023). “Measuring and Fighting for Price Stability in Turbulent Times: Lessons from East Asia”. In: *Inflation and Deflation in East Asia*. Ed. by F. Rövekamp et al. Cham: Springer Nature Switzerland, pp. 1–10.
- Solow, R. (1987). “We’d Better Watch Out”. In: *The New York Times Book Review*, 12 July, p. 36.
- U.S. Bureau of Economic Analysis (2026). *Real Gross Domestic Product per Capita [A939RX0Q048SBEA]*. Retrieved from FRED, Federal Reserve Bank of St. Louis. URL: <https://fred.stlouisfed.org/series/A939RX0Q048SBEA> (visited on 30 September 2026).
- Zenghelis, D. (2026). “Letter: Why the Demand Side Matters in Europe vs US”. In: *Financial Times*, 24 July.

We'd love to hear your feedback! Let us know below in our 30-second feedback survey what you enjoyed and what we can improve.

