

Geloso, Vincent; Reilly, Chandler S.

Article

National output without government? State capacity and welfare measurement

Journal of Government and Economics (JGE)

Provided in Cooperation with:

Elsevier

Suggested Citation: Geloso, Vincent; Reilly, Chandler S. (2025) : National output without government? State capacity and welfare measurement, Journal of Government and Economics (JGE), ISSN 2667-3193, Elsevier, Amsterdam, Vol. 19, pp. 1-14,
<https://doi.org/10.1016/j.jge.2025.100155>

This Version is available at:

<https://hdl.handle.net/10419/340768>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



National output without government? State capacity and welfare measurement

Vincent Geloso^{a,*,#}, Chandler S. Reilly^{b,#} 

^a George Mason University, Virginia, USA

^b Metropolitan State University of Denver, Denver, Colorado, USA

ARTICLE INFO

Keywords:

Economic growth

Private product remaining

State capacity

Gross domestic product

ABSTRACT

Should government services be counted in GDP? In this paper, we argue that this is the wrong question. The more relevant question is: what do government services allow us to capture about economic well-being? By construction, counting government spending at cost as part of GDP turns it into an upper-bound approximation of welfare—one that tends to overvalue the contribution of the state. We use the concept of the *Private Product Remaining* (PPR) from Rothbard (1972) as a lower-bound measure of economic output that removes government in a comprehensive manner from GDP. We argue that the gap between GDP and PPR reflects the uncertainty surrounding the true welfare contribution of the state (thus affecting the reliability of any attempts of using national accounts to speak to welfare). This gap becomes analytically useful once we introduce state capacity into the picture: improvements in state capacity shift us along the spectrum between these two bounds. We formalize this idea through a "measurement legitimacy" function of state capacity which lies between PPR and GDP depending on the effectiveness of the state. Using newly extended data for the United States from 1889 to 2024, we find that the size and direction of the gap between the two measures vary systematically in ways that alter our understanding of American economic history and the role of state capacity. From 1889 to 1928, rising state capacity leads to GDP understating growth. Between 1929 and 1985, GDP *overstates* growth. After 1985, GDP once again *understates* it.

1. Introduction

The construction of consistent national accounts stands as one of the crowning achievements of the economics profession. Simon Kuznets, the one at the forefront of the many pioneers of national accounts (Studenski, 1958; Fogel et al., 2013), was awarded the Nobel Prize for his foundational contributions. For most economists, any serious debate over what should or should not be included in GDP was settled long ago. As a compromise measure, GDP is acknowledged to have flaws—but these are well understood and accepted as the best among imperfect

alternatives. And GDP does a high-quality job in tracking other measures of living standards (Pritchett, 2022; Reinsdorf and Sheiner, 2024).¹

We respectfully depart from this consensus by scrutinizing a key component of GDP: the measurement of government's contribution to national output. Standard national accounting conventions treat most government spending as output. This approach risks overstating the economic value generated by the public sector. We propose an alternative framework that treats GDP as the upper bound of a conceptual range, with the lower bound defined by Murray Rothbard's (1972, 2004) notion of Private Product Remaining (PPR)—the portion of

* Corresponding author.

E-mail address: vgeloso@gmu.edu (V. Geloso).

Vincent and Chandler shared equal contribution for the project.

¹ Here, we are concerned with GDP as a measure of welfare. Scholars like Rockoff (1998, 2019) have argued that the relevance of GDP depends on the question being asked. For instance, in discussing the removal of wartime expenditures from GDP by Kuznets (1945) and Higgs (1992), Rockoff writes: "for other purposes, such as determining the pace of the mobilization, or comparing the performance of the United States with that of other belligerents, both central concerns of our present volume, an output measure that includes munitions is the only one that makes sense" (Rockoff, 1998, p. 84). In doing so, Rockoff implies that GDP-as-welfare differs from GDP-as-capacity—to stack tanks and guns against a national enemy. This may be somewhat overstated, as other national accounting systems—most notably the Soviet Material Product System (MPS)—were more directly suited to measure physical production for mobilization. Still, the spirit of his point is broadly correct, and this is why we emphasize that our concern lies squarely with GDP's use as a proxy for economic welfare.

output left in private hands after government appropriation.²

PPR was first presented by Rothbard as an alternative to broadly accepted measures of living standards, such as GDP, that removes all income originating in government and all government expenditures or receipts (whichever is higher). From Rothbard's perspective, his PPR brought us closer to what could be considered an accurate measure of living standards because it takes out all "government depredation." This conclusion is directly related to Rothbard's conception of welfare economics which led him to conclude that all government activity is wasteful because it is coercive. If what the government did was valuable, then it would already be happening in the market.

We do not share Rothbard's strong assumption that all government activity is inherently wasteful, but it is equally unreasonable to assume that all government activity is inherently valuable. This is particularly evident once we consider the incentives embedded in bureaucratic structures, which tend to drive up the cost of delivering any given public good or service (Niskanen, 1971; Migué and Bélanger, 1974; Tullock, 2005). The truth lies somewhere in between.

The main purpose of this paper is to show why PPR is a useful complementary (joint) measure of living standards, particularly in applied work aimed at determining the effects of the state on economic growth. GDP as the upper bound is output by a perfectly benevolent social planner (every dollar of spending has its social value). PPR as the lower bound is output in a society with a predatory ruler (every dollar of spending is a burden on the private economy). Situating government output within this range between PPR and GDP offers a more nuanced perspective on the economic role of the state and invites a reconsideration of how, and to what extent, government capacities contribute to development.

This distinction matters for any modern study of state involvement in the economy. Imagine a causal econometric test that measures the effect of a change in state capacity—defined as the state's ability to achieve its policy goals (Dincecco, 2017)—on economic activity. Using GDP as the measure of economic activity, one finds that state capacity increased activity. But that result reflects an embedded assumption: that state spending is entirely productive. If the same test is run using PPR instead of GDP, and the results still hold, then the effect is not driven by that assumption. If the results differ, it means the measurement method drives the outcome—and further work is needed to determine which of the measures GDP or PPR better reflects true economic activity.

We present a simple formal model to assess which metric—GDP or PPR—better approximates true economic activity. In the model, welfare is a function of PPR, GDP, and state capacity. This framework allows us to analyze how changes in state capacity affect welfare and to estimate the distance of each measure from the true level of welfare. By specifying assumptions about the "returns to state capacity," we can determine whether GDP or PPR provides a closer approximation to actual welfare as state capacity varies.

We simulate such changes under different assumptions of returns to state capacity to show that in most cases, GDP does indeed come closer to the true measure. But in a significant portion of cases PPR comes closer. We discuss the implications of this simulation for empirical research on the importance of the state.

We also construct and present the first long-run estimate of PPR for the United States. Earlier estimates of PPR in the U.S. have been limited in their coverage, primarily focusing on the 20th century. They were

also constructed using Gross National Product (GNP). To preserve comparability with these earlier estimates, we do the same but note that we can do the same with GDP with no loss of significance to our point.³ Using data from the Historical Statistics of the United States and the BEA NIPA tables, we calculate annual PPR from 1889–2024 for the United States. This series is used to ground our simulations of state capacity effects and the choice of living standards measure.

The rest of the paper is structured as follows. Section 2 reviews the debates over inclusion of government in national accounts and presents our case for treating PPR as a lower bound. Section 3 discusses the implications of treating PPR as a lower bound as it relates to the welfare effects of state capacity and presents our model. Section 4 describes the data we used to construct the long-run estimate of PPR in the U.S. from 1889–2024 and presents the series. After the series is presented, we illustrate the model using a state capacity index. Section 5 discusses and concludes.

2. National Accounts and The Contribution of Government

2.1. Valuation of Government Services: Between a Low and High Bound

When measured on the expenditures side, GDP and GNP were defined as the market value of all final goods and services produced within an economy over a given period, usually one year. This includes government consumption spending. This may include purchases of actual goods (e.g., pens and computers for employees) but it also includes the compensation of state employees (as it is a service).

The government account in GDP must impute value to what are considered outputs of non-market services, since these services are provided at zero prices (Lequiller and Blades, 2014, p. 280). Typically, non-market services—such as household production—are excluded from GDP. However, there are two main exceptions: (1) government services provided free of charge, and (2) housing services consumed by homeowners, imputed through ownership (Lequiller and Blades, 2014, p. 108). For the latter, statistical agencies estimate an "imputed rent," usually based on observable prices in the rental housing market. For the former, the task is more difficult: most government services are monopolized, and there are no comparable market prices to use as a benchmark.

Complicating matters further, governments often provide services well below cost, for a range of normative or policy-driven reasons. To estimate the value of these services in GDP, national accountants rely on the cost of production—that is, the state's actual expenditures. The imputed value of non-market government services is therefore calculated as the sum of: (1) compensation of employees, (2) purchases of goods and intermediate inputs, (3) consumption of fixed capital, and (4) payments for services provided to households (e.g., reimbursement of health care costs in publicly financed but privately provided systems). From this total, partial payments by households or firms for government services—such as museum entry fees or the purchase of government publications—are subtracted (Lequiller and Blades, 2014, p. 140).⁴ In other words, the value of government consumption is imputed based on the cost of production. This choice—now the norm and embedded in the definition most people think of when they think of GDP—was the subject of considerable debate among the pioneers of national accounting. Many acknowledged that including government output at cost overstated the size of the economy. Yet, it was viewed as a consistent and non-arbitrary

² PPR is different from private GDP where government spending is merely subtracted from total GDP.

³ Gross National Product (GNP) measures the total market value of all final goods and services produced by a country's residents, regardless of where the production occurs, while Gross Domestic Product (GDP) measures the value of all production that takes place within a country's borders, regardless of who produces it. The PPR estimate discussed in the main text is based on GNP, but we also include the results of a GDP-based PPR estimate in the Appendix.

⁴ See also Ruggles and Ruggles (1956, pp. 53–54).

shortcut, one with a known direction of bias—namely, that it would create an upper boundary for measuring economic activity.

Simon Kuznets, who won the Nobel in large part for his work creating national accounts and the methodology for them, "was interested in working out how to measure economic welfare rather than just output" (Aitken, 2019, p. R5). On that basis, he proposed multiple *lower* measures of output than the one conventionally used (Coyle, 2015). He suggested that expenditure on weapons advertising, and speculative financial activities should be subtracted from national accounts. He also proposed treating the government like a business, using tax receipts as a proxy for prices (Studenski, 1958, p. 195). In his view, taxes represented the price of government services.⁵ All of these attempts by Kuznets implies that he viewed "better" estimates of output as being smaller than the measured GDP.

In the end, the initial position of individuals like John Hicks (1940) who argued for no distinction and what is essentially the now-conventional treatment of government in GDP. However, as Spindler (1982) notes, they did so largely for "reasons of expediency" (p. 183).⁶ Overall, it was widely understood that this approach represented an upper boundary that overstates economic output. Alternative definitions were seen as potentially valid, depending on the specific question at hand.

This sense of expediency remains acknowledged today. Writing in the *Review of Income and Wealth*—the leading journal in national accounting—Bournay (2007) notes that "almost all government production is recorded as final demand (and totally included in GDP), despite the fact that part of it is intermediate consumption by the institutional sector," and concedes that this must lead to an overstatement of the economy's size (p. 742). Similarly, Zvi Griliches (1998), in his presidential address to the American Economic Association, described government as an "unmeasurable" sector and implied that its inclusion likely leads to overstatement. The Atkinson Review in the United Kingdom proposed alternative measures to better capture productivity growth in the public sector (Atkinson, 2005). The value-added approach recommended in that report suggested that measured levels of output would be lower relative to GDP, but that observed trends over time would be more accurate (ab Iorwerth, 2006). This implies that GDP is *known* to be an upper bound.

These issues are compounded by the challenge of converting nominal government spending into inflation-adjusted terms. As noted by Nordhaus (2002) and the Atkinson Review, there are missing prices for many public services. This makes conversion difficult, resulting in "defective deflators"—or at best, imperfect ones.⁷ Take healthcare provision as an example: because output is hard to measure directly, we rely on input costs (such as wages for support staff, nurses and doctors) as a proxy. However, if wages rise without a corresponding increase in the number

of doctors or patients treated, nominal GDP will increase even though real output has not. If deflators fail to adequately adjust for the input price changes in this example, it leads to an overstatement of real growth. In this way, poor deflators may not only cause level overvaluation, but also distort trend growth, compounding the problem.⁸ In such cases, we know that real GDP must overestimate – further cementing the point of it being an upper bound.

In other words, the conventional GDP measure is a consistent and less arbitrary metric than many alternatives, but it carries a known bias—namely, that it serves as an upper boundary for estimating economic output.

2.2. Private Product Remaining (PPR) as the Lower Boundary

What would be the lower boundary equivalent – in terms of consistency and predictability of the bias – to GDP? The best lower boundary measure is the Private Product Remaining (PPR). The main popularizer of the concept is Murray Rothbard (1972), who prominently used it in his study of the Great Depression and the role of government in causing it.⁹ He saw it as a "challenge to the orthodox postulate that government, *ipso facto*, represents a *net addition* to the national product" (p. 296). The "postulate"¹⁰ he criticized was that government spending bears a "necessary relation to the services it (the state) might be providing to the private sector." Instead, he argued it was far more "realistic to make the *opposite* assumption"—namely, that all government spending constitutes a "clear depredation (...) to private product and private output" (p. 296).

The key point is that Rothbard argued for relaxing the assumption of productiveness—an assumption implicit in the idea that the cost of government-provided services reflects their contribution to national product. He later contended that there is "no possible way to measure government's alleged 'productive contribution'" (Rothbard, 2004, p. 1293). This problem is compounded by the fact that government services are often monopolized, and "monopolized and inefficiently supplied" services, he argued, are worth less than their cost (Rothbard, 2004, p. 1293).

To calculate an alternative to conventional national accounts—the Private Product Remaining (PPR)—Rothbard started with GNP. This explains why, in the next section, we will shift to using GNP instead of GDP (allowing comparisons with his estimates and those of others). From GNP, one must first deduct "incomes originating in government" (i.e., the salaries of officials and employees of government enterprises). This yields what he called the "net private product," from which one then deducts the "depredations of government" to arrive at the final PPR measure (Rothbard, 2004, p. 1293). The depredations to be subtracted are defined as the greater of either government revenues or government expenditures. In other words, PPR can be calculated using the following procedure:

$$PPR = Y - AY_G - \max(AE, AR) \quad (1)$$

where E , R , Y_G and Y represent, respectively, expenditures, revenues, government-originated income and the usual Gross Domestic Product.

⁵ This approach, however, has since been discarded, based on the argument that "there is no direct link between the payment of a tax and the level of services required" (Lequiller and Blades, 2014, p. 114). Ruggles and Ruggles (1956), reflecting the consensus that had emerged by the 1950s, argued that using this would create a "fiction" under which tax payments "are equal to the amount at which government services are valued" (p. 52).

⁶ Hicks, alongside his wife Ursula, was pretty clear: How do we draw the line between the value of these services and the value of those services which ought to be deducted? The division seems to be entirely arbitrary. Consequently, if we want to measure something and not arrive at a figure for the national income which is what it is just because we say it is, it seems better to disregard this productive utilization of public services, and to regard them (by definition) as being reckoned entirely into final output (Hicks and Hicks, 1939, p. 150).

⁷ The Atkinson Review notes multiple reasons for this. One is that different government authorities produce inconsistently defined information (Atkinson, 2005, p. 61); poor data timeliness and periodicity (Atkinson, 2005, p. 64); poor capital input price data (Atkinson, 2005, p. 68); poor measurement of labour input prices (Atkinson, 2005, p. 70); poor weighting schemes (Atkinson, 2005, p. 72).

⁸ With respect to this specific example of doctor pay, the Atkinson Review mentioned multiple relevant examples. For example, deflators for labor services often relate to earnings only but not total employment costs (e.g., wages plus pension contributions, insurance premiums) and the mix of "types" of labor (e.g., skilled, unskilled) are not sufficiently disaggregated to allow to capture for changes in input mix that could be affecting costs but also output (Atkinson, 2005, p. 51).

⁹ He argues that the scale of government depredations in the early 1930s exceeded what conventional statistics reveal, thereby worsening the contractionary phase of the Great Depression from 1929 to 1933.

¹⁰ It is arguably unfair to call it a postulate. As noted above, it is more accurately described as an assumption made for reasons of expediency, non-arbitrariness and consistency.

The scalar $A > 0$ accounts for missing data on state and local government finances and is relevant only for the period 1889 to 1928, as discussed further below.¹¹ PPR is thus seen by Rothbard into a measure of true added value in the economy since the removal of depredations account "for the fact that such coerced expenditures/revenues are made at the cost of private expenditures" (Strow and Strow, 2013, p. 64).

It might be tempting to reject PPR as too extreme, given its assumption that government provides no value.¹² Indeed, Rothbard later elaborates on his view of the value of government output, which underpins his preference for PPR: "we must conclude that the government's productive contribution to the economy is precisely zero" (Rothbard, 2004, p. 940). He further argued that even conceding that "governmental services are worth *something* (...)" ignores the "unseen"—namely, what private individuals might have done instead in the absence of state intervention (Rothbard, 2004, p. 940).

At this point, many will likely part ways with Rothbard, perceiving that he goes too far—largely for two reasons.¹³ First, he conflates an accounting exercise (the valuation of current output) with a counterfactual scenario in which the economy operates without government, which he assumes would yield a higher level of output. It may very well be true that output would be larger without government, but in that case, *both* PPR and conventional GDP would be higher. This is qualitatively different from asserting that government output has zero value to consumers.¹⁴ In other words, Rothbard confuses his anarcho-capitalist normative viewpoint with measurement. The measurement does not require the normative viewpoint. Second, Rothbard assumes that government generates no positive spillovers and provides no public goods (Rothbard, 2004, pp. 1029–1041). To be fair, he later offers a justification for this assumption, but few within the broader economics profession are likely to accept it.¹⁵

Whether all this amounts to an extreme position is, for our purposes, irrelevant. The charge of being "too extreme" (or not extreme enough) does not, by itself, invalidate an idea's analytical usefulness. In fact, Rothbard himself made a subtle concession that implicitly softens his stance. In doing so, he acknowledges that the Private Product Remaining (PPR) is best understood not as a precise estimate, but as a conceptual lower bound for assessing government's contribution to output. As he put it: "any person who believes that there is more than 50% waste in government will have to grant that our assumption is more realistic than the standard one" (Rothbard, 1972, p. 296). The value of PPR lies in providing the lower bound of an interval, with conventional GDP as the upper bound and the true measure of economic activity located

somewhere in between. The lower bound represented by PPR assumes that the state's actions are *purely* extractive and contribute no value. The upper bound, by contrast, treats the state as an "optimal" provider—one that maximizes social welfare without regard to the utility or preferences of the planner.

Other works that rely on close cousins of PPR also accept the idea of boundary measures, though framed slightly differently. One such cousin is the Restricted Market Production Concept (RMPC). Developed by Polish Marxist economists working on national accounts, RMPC measures activity that passes through the market—though not necessarily private activity, since some government services, such as alcohol sales in state-run stores, occur in competitive markets.¹⁶ The underlying justification is that market transactions reveal value (Kalecki and Landau, 1934; Landau, 1934). It can be calculated by treating payments by the government to employees as transfer payments which means that they are excluded from GDP (Spindler, 1982). This is essentially a tweak on the PPR equation since $RMPC = Y - AY_G$ (i.e., there are no removal of Rothbard's depredations). Spindler (1982) resurrected the RMPC by realizing that that it provided a measure relevant to key points regarding public choice theory and the nature of bureaucracies. Relying on Niskanen (1971) and Migué and Bélanger (1974), he points that bureaucracies have incentives to inflate costs or operate inefficiently, which distorts the value of their output when measured using standard GDP accounting.¹⁷ In other words, costs are inflated due to bureaucratic behavior (budget maximization, slack, inefficiencies). By definition, there must be *some* deviation from the upper boundary. The deviation could be as large as total if *all* of the spending is waste.¹⁸ How much of a deviation is something Spindler (1982) refuses to commit to which is why he presents RMPC and the conventional measures as providing a range.

Nobel laureate James Buchanan and his co-author Francesco Forte also developed an alternative measure that accepts the idea of boundaries (Forte and Buchanan, 1961). Their argument is that, to the extent government contributes to the economy through productive activities (e.g., the provision of public goods), this value will be reflected in the market value of privately produced outputs. Therefore, government should not be included directly, and only private output should be measured. This measure is denoted PY, or private GDP. If could stimulate the economy—even if the activity produced nothing of actual value, government spending on productive activities is fully appropriated by private actors and thus entirely priced into market transactions, the lower-bound estimate proposed by Forte and Buchanan (1961) would equal the true measure of economic activity—denoted TPY for true

¹¹ After 1928, $A = 1$.

¹² Batemarco (1987) provided an estimate of PPR for the 1947 to 1983 period and recognized that many of those that do "not sharing Rothbard's anarcho-capitalist leanings... would recoil from the assumption that the government produces nothing of value" (1987, p. 185). We are part of those who recoil.

¹³ There is a third one – ideological but we emphasize the other two because we want the criticism that would speak to everyone independent of ideology. However, Evans (2018) summarizes this criticism best when he states that 'the real meat of the Austrian justification for PPR seems to be based on libertarian, not economic reasoning. If you believe that tax is theft, then indeed it seems appropriate to count government spending as categorically different to voluntary spending' (p. 133).

¹⁴ In fact, in his estimation of PPR from 1947 to 1983, Batemarco (1987) implicitly acknowledges this point. He cites evidence suggesting that people derive *some* value from government spending, in proportions indicating that less than 50% of it is perceived as waste. Under such conditions, conventional GDP would be closer to the true measure of economic activity than PPR. A similar point is made by Evans (2018, p. 134).

¹⁵ Batemarco, who produced the first long-run estimate of PPR (extending beyond just the Great Depression, unlike Rothbard), also advanced arguments in support of Rothbard's rejection of conventional output measures. He argued that most negative "spillovers" are generated by the state, while most positive externalities are byproducts of private activity (Batemarco, 2023).

¹⁶ It is worth noting that the Soviet system of national accounts – the Material Product System (MPS) – implicitly made the same distinction. MPS excludes most services, treating them as non-productive, and emphasizes gross output rather than value added. As most government activities are seen as services, they are non-productive and thus excluded from national output measures. Only government activities that directly contributed to material production through state enterprises were included.

¹⁷ In Niskanen (1971), bureaucracies are modeled as monopolistic suppliers of goods that seek to maximize their budgets rather than social welfare. Their patrons—politicians—have incentives to acquiesce to higher production costs. First, because politicians respond to political (i.e., non-market) incentives such as re-election odds and satisfying interest groups (i.e., rent-seeking). Second, because bureaucrats have a comparative advantage in understanding the information of the bureaucracy, they can control the flow of information to the politician. They can strategically withhold or frame information in ways that they can get politicians to acquiesce to the higher production costs. This is because what bureaucracies do is technically complex or opaque – so much that politicians and taxpayers have a hard time evaluating, auditing and sanctioning. As a result, politicians acquiesce. The outcome is the costlier provision of services for any given quantity of a public good.

¹⁸ An extreme example is the common analogy directed at early Keynesians: that digging and filling holes

private GDP. However, if appropriation is imperfect, their measure PY will understate the true level of economic activity. Implicitly, Forte and Buchanan (1961) accept the boundary logic, but they express it as a weak inequality in the case of full appropriation: $PY \leq TPY < Y$.

Somewhere between Spindler (1982) and Forte and Buchanan (1961), there are others—such as Robert Higgs (1992), Kenneth Boulding (1993), Simon Kuznets (1945), and James Tobin and William Nordhaus (1972)—who argue for excluding defense spending, or at least large portions of it. Their reasoning aligns closely with that of Forte and Buchanan (1961), but they add an important nuance: “defense expenditures have no direct value in household consumption,” and present a valuation problem since “no reasonable nation purchases [defense] because its services are desired *per se*” (Nordhaus and Tobin, 1972, p. 28). Boulding summarized a position closer to that of Spindler (1982), arguing that the assumption that “the product of the war industry is equal to its cost” was a “dubious assumption” (Boulding, 1993, p. 27). Geloso and Reilly (2025) constructed a measure reflecting these arguments about defense spending and argued that it represents a value somewhere between the true measure of economic activity and the overstated one. In other words, they accept the idea of a bounded range, with the lower boundary defined by the PPR (and their defense-adjusted estimate being within the range).

Finally, there is also Randall Holcombe (2004), who comes closest to endorsing PPR without fully adopting it. He argued that we should exclude “most government output” because it largely consists of intermediate goods and is typically given away rather than sold. The first reason aligns with the argument of Forte and Buchanan (1961) and need not be repeated here. The second reason draws on the logic of what is normally excluded from conventional GDP. As Holcombe notes, “much final output in an economy has value but is excluded from GDP because it is not sold,” citing domestic non-market production as an example (Holcombe, 2004, p. 395). “Highways may have value,” he continues, “but like personal home repair projects and home cooking, that output is not sold on the market” (Holcombe, 2004, p. 395). Thus, “for consistency,” government outputs that are not exchanged in markets—such as policing, fire services, or roads—should also be excluded from GDP (Holcombe, 2004, p. 395).¹⁹ However, under Holcombe’s definition, the sale of alcohol by state-owned corporations (e.g., Virginia’s Alcoholic Beverage Control Authority), the sale of electricity by state-owned utilities (as in the case of Canadian provincial power companies), or—more historically—fares collected by state-owned airlines, would all be included in GDP, since they involve market transactions.²⁰ These are a wide range of alternative measures that are produced from economists across the entire spectrum of economic schools of thought. However, all of these alternative measures should be understood as falling within the interval defined by Private Product Remaining and Conventional GDP. Each represents a distinct accounting exercise aimed at producing reliable estimates within a clearly bounded range. Each is larger than PPR and smaller than conventional GDP. Which one is correct is irrelevant here. What matters is not even PPR in isolation, but PPR together with its counterpart at the upper end of the interval (i.e. conventional GDP). PPR is simply the most methodologically consistent and conservative lower bound. By allowing us to gauge the size of that interval, PPR serves as a crucial tool in the practice of economic measurement.

This is particularly relevant, as we will argue below, to several empirical debates in economics. The “quality” of government interventions determines where within the interval the true measure of economic activity lies. Imagine an ideal world governed by a benevolent

social planner—this would imply that the upper bound (i.e., conventional GDP) reflects the true value. Now imagine a dystopian world ruled by a predatory state bent on extraction—here, the lower bound (i.e., PPR) would be the accurate measure. A shift from the latter to the former implies an upward move within the boundary; the reverse implies a downward move. Relying on a single measure to capture such shifts in the real world will inevitably introduce bias. Using both measures—that is, both boundaries—helps attenuate some of that bias.²¹

3. Implications

Our framing of the joint use of PPR and conventional GDP as boundaries on a range of values regarding government’s contribution to the economy is rich with implications. They all fall under the rubric of “robustness”. Consider our key contention: it affects our understanding of the role of state capacity in economic development. Using PPR and conventional GDP will help identify the bias of any empirical study relating state capacity to development.

First, it is necessary to clarify the concept of state capacity and explain how its relationship to economic development should be understood through our framework of measurement using both PPR and GDP. State capacity is commonly defined as “the government’s ability to accomplish its intended policy goals” (Dincecco, 2017, p. 3). These goals can be grouped into three broad categories that define ideal-types of states: the protective, the productive, and the redistributive/predatory (Buchanan, 1975).

The protective state is limited to enforcing agreed-upon rights, such as property rights. It can be likened to the analogy of the nightwatchman state via courts, policing and national defense. The productive state does what the protective state does but adds market-enhancing functions. These include the standard justifications found in public economics: regulating natural monopolies, dealing with externalities, producing public goods, and delivering education (see Li and Maskin (2021) for a survey). If governments possess the capacity to enforce formal rules under the protective framework, they may also support market-enhancing institutions, thereby qualifying as productive states (Besley and Persson, 2008, 2009, 2010; Dincecco and Katz, 2014; Dincecco, 2015, 2017).²² For many of the state capacity theorists, the productive state is superior to the protective state in terms of outcomes (e.g. Hanlon (2024)).

To see this normative superiority, the central concept is the “encompassing interest” of rulers. If rulers benefit from a share of overall output, they are incentivized to invest in public services that expand the size of the economic pie—even if their share remains constant. Constraints (in the form of possible rebellions, democratic turnover, inter-jurisdictional competition, and the ability of people and capital to move) are sticks that complement incentives of the carrot of the share of output that they get (Olson, 1993; Johnson and Koyama, 2017; Acemoglu et al., 2015; North et al., 2009). In other words, expanding markets enhances the ruler’s wealth.²³

The redistributive state, for its part either substitutes itself for the private sector or attempts to control it (Vahabi, 2016; Buchanan, 1975; Anderson and Hill, 1989; Geloso and Salter, 2020). Its activities revolve primarily around transfers via rent-seeking rather than enhancing markets. Redistribution can occur in both democracies and

²¹ We say “some” because the ideal would be a real-valued function that precisely locates both the start and end points within the interval.

²² For comprehensive literature reviews on state capacity, see Cingolani (2013); Savoia and Sen (2015); Bardhan (2016), Koyama (2016), Johnson and Koyama (2017), and Piano (2019).

²³ Good historical examples of this are the cases of women’s property rights in pre-1914 America (Lemke (2016)), railways in Prussia (Uebele and Gallardo-Albarrán, 2015), and post office in America and Canada (Acemoglu et al., 2016; Aneja and Xu, 2022; Geloso and Makovi, 2022).

¹⁹ However, it is worth noting that what does not get into GDP can still be measured. These are generally measured as parts of “satellite accounts” such as those for illicit economy (Goel et al., 2019; Soloveichik, 2019) and home production (Wagman and Folbre, 1996; Landefeld et al., 2009).

²⁰ PPR – since these are incomes originating in government – would exclude them.

dictatorships, but in the latter, the term “predatory state” is more appropriate. A predatory state prioritizes the enrichment of rulers or elites over the well-being of the general population. Yet, much like its democratic analogue, its essential nature lies in reallocating wealth through politically driven, rent-seeking mechanisms. For simplicity of exposition here, we can assume that the Hobbesian jungle is the same as a predatory state (see footnote 24 for more details on that assumption). State capacity should not be interpreted to refer simply to a “bigger state,” but to a state that can better enable economic activity. This can potentially involve a larger state. But it can also involve a state that reduces its costs without decreasing the quantity or quality of its services. The creation of a new public good (i.e., a cost increase) is the conceptual equivalent to a reduction in the cost of government services provided output and quality remain the same.

With respect to measurement, an increase in state capacity corresponds to a strengthening of either the protective or the productive functions of the state. These are moves *away* from the predatory/redistributive state. According to state capacity theorists any increase in state capacity enhances the productivity of private economic activity.²⁴ The central issue, therefore, is how to identify the appropriate metric (PPR or GDP) for assessing the effects of changes in state capacity. When the state is predominantly predatory, PPR offers a more accurate reflection of economic reality since 100% of the state’s expenditures are for unproductive purposes. If there is a modest shift away from predation through the buildup of state capacity, PPR must rise as less predatory behaviour must boost private sector productivity. However, the relevance of GDP should also increase because *some* of the state’s costs (which enters GDP) can now be read as truly welfare-enhancing. However, because increases in state capacity can be associated with either smaller or larger governments, the gap between PPR and GDP may either narrow or widen. This leaves the interpretation of the effects—and especially their magnitude—ambiguous when relying on a single measure.

To see this, let us formalize the reasoning. Let $X \in [0, 1]$ and $X \in \mathbb{R}$ represent what the state does, where $X = 0$ corresponds to a fully predatory government and $X = 1$ corresponds to a fully benevolent government. In other words, this is a world where one goes from extractive to productive state. Let PPR denote the Private Product Remaining, representing the lower bound of measured economic activity, and let GDP denote Gross Domestic Product, representing the upper bound. PPR is the true measure of economic activity if the state is fully predatory. GDP is the true measure of economic activity if the state is benevolent and productive. We now define the true national account approximation of welfare, $Z(X)$, as a combination of PPR and GNP:

$$Z(X) = f(X) \begin{cases} f(0) = \text{PPR} \\ f(1) = \text{GDP} \end{cases} \quad (2)$$

²⁴ We do not fully share this view, though we acknowledge it as the dominant position in the literature. An alternative perspective can be referred as the “governance view.” This view can be summarized by the idea that “the production of rules is itself a productive activity,” meaning that governance is treated as a good that must be produced like any other. Within this framework, the state is considered only one possible provider of governance, rather than its sole or necessary source (Taylor, 1976, 1982; Rothbard, 1978; Mueller, 1988; Benson, 1990; Ellickson, 1994; De Jasay, 1998; Leeson, 2014; Stringham, 2015; Piano and Rouanet, 2020; Hasnas, 2024). The primary advantage of the governance view is that it allows us to remain agnostic regarding the feasibility or long-term stability of stateless orders (Geloso and Salter, 2020). While we think this view is analytically advantageous, our point about measurement can be encompassed within it. Under statelessness, all governance would, by definition, be provided outside of government, in which case Private Product Remaining (PPR) would equal Gross Domestic Product (GDP). Thus, our point about measurement is not orthogonal to the governance view, but rather closely aligned with it in how it conceptualizes the production and measurement of rule-making activity.

A flexible and interpretable form for capturing the relationship between state capacity and economic outcomes is a power function:

$$Z(X) = \text{PPR} + (\text{GDP} - \text{PPR}) \cdot X^\gamma \quad (3)$$

In this specification, $\gamma = 1$ yields a *linear* (weakly convex) relationship, $\gamma > 1$ implies a *strictly convex* form—suggesting *slow initial gains* from increases in state capacity, X , while $\gamma < 1$ yields a *concave* shape—implying *larger early gains* from initial improvements in X . This formulation allows state capacity to determine the position between PPR and GDP, depending on the shape of the function that maps state capacity into economic performance. Now, imagine that we want to know the effect of state capacity on living standards more broadly. A change in X will have two channels: one acting directly on both GDP and PPR, and another affecting Z indirectly via changes in the relationship between the two. This means that state capacity may raise both measures, but also shift their relative importance, as a “better” state makes GDP relatively more informative about well-being than PPR was before the change in X . The *true* welfare (W) from this change in state capacity can be denoted $W(X, Z(X))$. The purpose of inserting $Z(X)$ into the W is to reflect how state capacity legitimizes the upper bound (GDP) as a better measure, in addition to any effect on productivity not reflected in the measurement legitimacy shift. In other words, the change in welfare can be broken down as follows:

$$\Delta W = \underbrace{\phi \cdot \Delta X}_{\text{Productive Effect of State Capacity}} + \underbrace{\psi \cdot \Delta Z(X)}_{\text{Measurement Legitimacy Shift}} \quad (4)$$

Under the logic of some state capacity theorists and more traditional public finance theorists, the state’s productive activities (e.g., solving market failures or engaging in market-enhancing interventions) directly improve welfare, which would increase both PPR and GDP. However, that effect may be uneven across the two measures.

Consider the example of a lighthouse—the ultimate textbook example of a public good. Providing a lighthouse will here be treated as analogous to providing more state capacity (a change in X). Once a lighthouse is provided, it boosts both PPR and GDP. Moreover, because a single lighthouse provides the entire market without rivalry in consumption, the size of PPR and GDP does not affect cost. For simplicity and illustration, we will assume an annual cost of \$10. Imagine now that $\text{PPR} = \theta$ and $\text{GDP} = \alpha \cdot \theta$ where $\alpha > 1$. The lighthouse increases PPR a factor of 1.5. Ergo, $\text{PPR} = 1.5 \cdot \theta$. Since the cost of the lighthouse is extracted from

GDP to arrive at PPR, it has only the productivity effect of the state capacity ($\phi \cdot \Delta X = 1.5$). However, GDP is now equal to $\text{GDP} = 1.5 \cdot \alpha \cdot \theta + 10$. This makes it evident that a state capacity improvement of this type increases PPR and GDP unevenly. This changes the parameters in $Z(X)$ which captures the fact that the state providing the lighthouse increases the relative relevance of GDP as a measure.

Now, imagine an alternative case of change in X which is that the state is now capable of providing the lighthouse (or whichever public good one thinks of) but using fewer taxes, fewer government employees, fewer bureaucratic steps and fewer resources. One useful example is how greater state capacity—through improved oversight of procurement and public employees—can enhance government efficiency. In low-capacity settings, procurement processes, even when formalized through tendering, may be used to dispense patronage, leading to inflated costs. Similarly, when employee performance is difficult to monitor, the result may be lower-quality services, excessive wages, or overstaffing.²⁵ In contrast, high-capacity states often achieve better

²⁵ A good example is provided by Romero (2025). More capable bureaucrats that are harder to evaluate because of their technical skills may use them to obscure favoritism rather than reduce it (leading to higher cases). Using more than 50,000 Guatemalan municipal contracts, he found procurement officers were effective at shielding politically connected firms by manipulating tenders in harder-to-detect ways.

value through transparent procurement and effective supervision, allowing the same nominal spending to produce greater real output.²⁶ In those cases, one could visualize that the same level of public services is being provided but for lower costs.

Thus, imagine that the size of the state falls by \$10 to reflect more efficiency in lighthouse provision. Assuming the same direct effect as in the example above, the new PPR is higher because of lower taxes, greater productivity, better allocation, fewer depredations etc. GDP will have to increase by at least the amount of PPR (since the lower boundary is part of the higher boundary). This means that, keeping same proportions as in the first lighthouse example, $PPR = 1.5 \cdot \theta + 10$. However, GDP must rise by less since $GNP = 1.5 \cdot \alpha \cdot \theta - 10$. Again, the rise is uneven for both measures but this time, PPR increases proportionally more. This gives us everything we need to understand the potential biases in any empirical setup attempting to link state capacity to economic welfare measures. The bias will depend on the type of state capacity improvement because the type affects the size of the gap between PPR and GDP as well the shape of the "legitimacy of GDP" effect (captured by γ in Eq. 3).

If one *forcibly* uses either GDP or PPR, the biases will work differently in getting the *true* effect on welfare (W).

Below, we simulate an illustration of these biases under different scenarios. We set $\gamma = 0.5$ or 1.5 to reflect concave and convex cases, respectively. Initial values are set as $PPR = 100$ and $GDP = 1.5 \cdot PPR$. We assume that a change in state capacity leads to a 50% increase in PPR, such that the new $GDP = 1.5 \cdot (1.5 \cdot PPR) \pm 10$. The $\pm$ term captures the two types of state capacity improvements: either a larger state (with higher fiscal inputs) or a more efficient state (delivering the same output with fewer inputs). We simulate changes of

$\Delta X = 0.1$, starting from either $X = 0.25$ or $X = 0.75$. We then compute the implied increase in true welfare and compare which of the two measures—PPR or GDP—is closer to the welfare- implied value. Table 1 below illustrates the results of this simulation. As can be seen, in some cases PPR provides a closer approximation of welfare than GDP, particularly when state capacity improves through reductions in the size of government without compromising output or quality. However, when state capacity increases through larger government provision, GDP consistently offers the closer approximation.

In the context of the state capacity literature, this issue is of considerable importance. Imagine that an empirical study—such as Knutsen (2013)—uses GDP per capita as the sole measure to estimate the effect of changes in state capacity. If the sample includes a significant number of predatory states, then relying on GDP will misstate the true changes in well-being. Moreover, if improvements in state capacity are achieved through a reduction in the size of government (e.g., lower taxes or leaner bureaucracies), GDP will again provide a biased representation of the effect. The magnitude and direction of these errors depend crucially on the true nature of the government's contribution to the *measurement legitimacy shift*, as captured by the parameter γ . The net effect risks severely biasing point estimates but also inflates standard errors, making statistical inference more difficult and potentially obscuring the true relationship between state capacity and development. Similarly, focusing only on PPR will miss the relevance of state capacity changes entirely by considering, automatically that all of government's activities are worthless.

We should also note that the measurement legitimacy shift offers a low-cost roundabout alternative to the data-intensive approach proposed in the Atkinson Review for improving the valuation of government services. The Review recommends several steps to mitigate trend

mis-estimation – particularly in response to issues like defective deflators and Griliches' (1998) concern with the "unmeasurable" sector. While these steps are relatively straightforward to implement in countries with great state capacity, they present serious difficulties in settings with weaker state capacities. When applied inconsistently across countries, steps such as those in the Atkinson Review risk introducing artificial differences in levels and trends due to methodological variation, thereby undermining international comparisons across large cross-sections or panel datasets. In contrast, our proposed measurement legitimacy shift circumvents this problem. As state capacity improves, allowing for more accurate data and better valuation of GDP, the legitimacy of GDP as a welfare indicator rises—naturally shifting measurement weight toward it without imposing inconsistent valuation methods across countries.

This already offers a valuable insight for applied economists: our approach, which treats measurement legitimacy as a function of state capacity, can help assess whether key empirical results are influenced by the choice to use GDP—potentially introducing uneven biases in favor of or against certain conclusions. In particular, specifications grounded in Eq. (4) provide a useful robustness check when studying the welfare effects of variables like trade openness or ethno-linguistic fractionalization or even of results such as the environmental Kuznets curve and the inequality Kuznets curve (both posit a quadratic relationship with income). The reliance on the measurement legitimacy shift approach would allow researchers to determine whether observed effects are genuine or merely artifacts of the underlying measurement framework.

It also offers another roundabout solution to studies of state capacity. One key aspect of state capacity improvements is the variety of public services that governments produce. Governments are often accused of being inflexible in production methods and thus they apply one-size-fits-all public goods and services. However, state capacity can allow some variety in those goods and services – something of value to consumers that a simple sum of expenditures may fail to reveal. For instance, Feenstra et al. (2020) show that Chinese cities with larger populations offer more consumer variety, leading to lower measured prices and higher real consumption. A high-capacity state may not only deliver services more efficiently, but also provide a more diverse set of public goods, externalities management (e.g., pollution control could be made more flexible by region or industry), regulatory frameworks, or legal protections. If that logic holds, then a higher-capacity state will provide a more varied set of goods and services at lower "prices." As discussed, there are no market prices for many state-provided goods and services – thus the value of variety may be missed because of defective deflators.

However, our approach circumvents this. With greater state capacity, there is greater variety in state-provided goods and services which would make GDP the more legitimate measure as greater variety is an indication of lower prices.²⁷ If expenditures fall while variety increases because of higher state capacity, the measurement legitimacy shift is towards PPR. If expenditures increase while variety increases, the measurement legitimacy shift is towards GDP. Just as in the case of the deflators, our approach offers a simple solution – via the distance between GDP and PPR – to assessing the true effect of state capacity changes. In other words, our model can subsume a large set of issues concerning how to properly evaluate the effect of governments in the economy.

4. Data, Results, and Illustrations

How can we illustrate the importance of this core point? One way is to construct the PPR measure and compare it with GDP for a single country over a long historical period. This allows us to assess how the

²⁶ For example, government audits conducted by external agencies have been found to reduce corruption and inflated procurement costs (Olken, 2007; Avis et al., 2018). Similarly, state capacity investments could also send strong signals to financial markets in ways that reduce borrowing costs (and thus expenditures on debt servicing) (Nomo Beyala, 2025).

²⁷ Our assumption here is that if increases in state capacity, X , lead to increases in variety, then the increases in X will capture the increases in variety limiting the need for a direct measure of variety.

Table 1

Simulation of state capacity effects on PPR, GDP, and legitimacy-weighted welfare.

PPR_1	PPR_2	$GDP_1 = 1.5 \cdot PPR_1$	$GDP_2 = (1.5 \cdot PPR_2) \pm 10$	Initial X	ΔX	γ	$\Delta Z(\%)$	$\Delta GDP(\%)$	$\Delta PPR(\%)$	Closer?
100	150	150	235	0.25	0.1	0.5	60.2	56.7	50.0	GDP
100	150	150	235	0.75	0.1	0.5	59.4	56.7	50.0	GDP
100	150	150	235	0.25	0.1	1.5	57.7	56.7	50.0	GDP
100	150	150	235	0.75	0.1	1.5	63.5	56.7	50.0	GDP
100	150	150	215	0.25	0.1	0.5	50.8	43.3	50.0	PPR
100	150	150	215	0.75	0.1	0.5	45.5	43.3	50.0	GDP
100	150	150	215	0.25	0.1	1.5	53.8	43.3	50.0	PPR
100	150	150	215	0.75	0.1	1.5	51.7	43.3	50.0	PPR

relationship between private retention and total output evolves over time. We apply a measure of state capacity— X , as defined in the previous section—to simulate $Z(X)$ and determine where, along the continuum between PPR and GDP, the best approximation of true welfare lies. To implement this approach, we use data for the United States from 1889 to the present. Because we also aim to compare our PPR series with previously constructed series for the periods 1929–1932 and 1947–1983 (Rothbard, 1972; Batemarco, 1987), which were based on gross national product, we use GNP as our primary reference. However, for comprehensiveness, we also compute PPR using GDP as the base, and we report these results in the appendix.

4.1. Data

4.1.1. Data for 1889–1928

We collect data on national income, government employment, government income, and government finances for the years 1889–1928 from the *Historical Statistics of the United States* (U.S. Census Bureau, 1975). Gross National Product (GNP) is taken from Series F-1, and the GNP implicit price deflator is from Series F-5 (U.S. Census Bureau, 1975).²⁸

Calculating PPR requires subtracting income originating in government, but the data for this period is incomplete. However, we have sufficient data to construct a plausible estimate. We first collect data on total civilian employment in the federal government.²⁹ This series shows federal government employment figures from 1816–1970 with missing data for years 1884–1890, 1892–1900, and 1902–1907. To interpolate missing values, we calculate the compound annual growth rate between the last available year before a gap and the first available year after. For example, to estimate employment for 1884–1890, we compute the annualized growth rate between 1883 and 1891 and assume constant growth across the intervening years.

Next, we collect data on average annual earnings for federal employees in executive departments available for the years 1892–1926.³⁰ To estimate missing values for 1889–1891 and 1927–1928, we apply the same interpolation method described above, using compound annual growth rates. Specifically, we backcast values for 1889–1891 based on the growth rate from 1892–1897, and forecast values for 1927–1928 using the rate from 1921–1926. Annual income for federal civilian employees from 1889 to 1928 is then calculated as the product of the employment and average earning series.

We supplement income originating in government for 1889–1928 with data on active duty military personnel and military pay.³¹ The personnel data are complete for all years in this period. Totals are

calculated separately for enlisted personnel and officers by summing across the three branches for each year.

Data on military pay are much more limited for this period in U.S. Census Bureau (1975). Weighted averages of annual pay for enlisted personnel and officers are available only for 1865, 1898, and 1918. We interpolate missing values using the compound annual growth rate between available data points.³² Total annual income for enlisted personnel is calculated by multiplying their average annual pay by the number of enlisted personnel in each year. Officer income is calculated analogously. These two series are then combined with the income series for civilian federal employees to construct total income originating in government for 1889–1928. We note that this measure includes only federal employees; the treatment of state and local government is addressed below in the discussion of government expenditure and receipt data.

To compute PPR, we subtract from GPP the greater of federal government expenditures or receipts. Fortunately, U.S. Census Bureau (1975) provides data on federal government expenditures and receipts dating back to 1792.³³ We collect both series and keep the maximum of the two for each year.

For the years 1889 to 1928, the *Historical Statistics of the United States* primarily provide federal government data and lack sufficient detail on state and local governments. To address this, we rely on the data compiled by John Wallis (2000), which report the revenue-side size of federal, state, and local governments from 1800 onward. As Wallis notes, the federal government was historically smaller than the combined size of state and local governments—a pattern that persisted until the post-WWII era. To incorporate state and local governments for the period up to 1929, we define an "augmentation ratio," in year $t < 1929$ denoted by A_t , as:

$$A = \frac{G_{A,t}}{G_{f,t}}$$

where G_f is the size of the federal government and G_A is the aggregate size of all levels of government in year t . For the 1889–1928, all estimated government items subtracted from GNP are multiplied by A_t . The resulting augmentation factors are shown in Table 2. The augmentation factors range from just over 2 to around 3, implying that total government activity was typically 2 to 3 times the size of federal government alone during this period.

4.1.2. Data for 1929–2024

Data on national income, government income, and government finances from 1929 to 2024 are more complete and require fewer assumptions. GNP for this period is obtained from the Bureau of Economic Analysis (BEA) accessed through FRED (U.S. Bureau of Economic Analysis, 2025). However, to ensure consistency with the pre-1929 data, we apply a modification following the procedure in Geloso and Reilly

²⁸ This price deflator is set such that 1958 = 100. However, when combined with the GNP price deflator for 1929–2024, it is reset to 2017 = 100 and spliced according to the procedure described in the section that follows.

²⁹ Series Y-308 in U.S. Census Bureau (1975).

³⁰ Series D-790 in U.S. Census Bureau (1975).

³¹ Data cover active duty personnel in the Army, Navy, and Marines (U.S. Census Bureau, 1975). Army enlisted and officers are Series Y-907 and Y-906, respectively. Navy enlisted and officers are Series Y-913 and Y-912. Marines enlisted and officers are Series Y-916 and Y-915.

³² The growth rate from 1865 to 1898 is used to interpolate average annual pay for 1889–1897; the rate from 1898 to 1918 is used to estimate values for 1899–1917 and 1919–1928.

³³ Series Y-335 and Y-336 in U.S. Census Bureau (1975).

Table 2

Government augmentation factor, 1889–1928.

Year	Factor	Year	Factor	Year	Factor
1889	2.33	1904	2.70	1919	2.49
1890	2.36	1905	2.74	1920	2.38
1891	2.40	1906	2.79	1921	2.28
1892	2.44	1907	2.84	1922	2.17
1893	2.48	1908	2.89	1923	2.28
1894	2.52	1909	2.93	1924	2.39
1895	2.56	1910	2.98	1925	2.50
1896	2.60	1911	3.03	1926	2.61
1897	2.64	1912	3.08	1927	2.72
1898	2.68	1913	3.13	1928	2.71
1899	2.71	1914	3.02		
1900	2.75	1915	2.91		
1901	2.68	1916	2.81		
1902	2.60	1917	2.70		
1903	2.65	1918	2.60		

(2025). Specifically, we splice the [U.S. Census Bureau \(1975\)](#) GNP series with the BEA series using two steps: (1) index the [U.S. Census Bureau \(1975\)](#) GNP series so that 1929 = 1, and (2) multiply these indexed values by the 1929 GNP level from the BEA series. This procedure preserves the movements of the 1889–1928 series while matching the level of the BEA series. The BEA implicit price deflator for GNP is also accessed via FRED and spliced with the price deflator from [U.S. Census Bureau \(1975\)](#) using a similar procedure, with all values re-indexed such that 2017 = 100.

Income originating in government for 1929–2024 is collected from the BEA National Income and Products Accounts (NIPA), specifically Table 2.1 on personal income by industry. We use personal income data attributed to the government sector, which includes income from federal, state, and local governments. Government expenditure and receipts for all levels of government are taken from NIPA Table 3.1; as before, we retain the maximum of the two for each year.

Population data for all years 1889–2024 are sourced from the Measuring Worth database and used to construct per capita measures ([Johnston and Williamson, 2025](#)). The pre-1929 data described in the previous section are combined with the 1929–2024 data to form the complete series used to calculate PPR for the full 1889–2024 period.

4.2. Results

We begin by presenting the complete constructed PPR series alongside GNP, as shown in [Fig. 1](#). From the start of the series in 1889 through the early 1940s, GNP and PPR closely track each other. While PPR remains consistently below GNP (as expected by construction) the two series generally move together. One notable exception to this early co-movement occurs during World War I. However, after World War II, GNP and PPR begin to diverge sharply, with GNP rising more steeply than PPR. Most of the divergence starts in WW2 and its aftermath.

This can be easily explained by the state growing larger. [Table 3](#) presents the average share of government depredation relative to GNP across selected historical periods. From 1889 to 1928, depredation averaged just under 20 percent, indicating that government activity—while modest by modern standards—still demanded a substantial portion of the private economy. This figure increases markedly during the 1929–1945 period, reflecting the fiscal expansions of the New Deal and World War II. The upward trend then continues in the postwar decades. The value of the PPR metric becomes especially clear here: unlike a simple ratio of government spending to GNP, PPR subtracts both the income originating in government (capturing its “productive” component) and the larger of government expenditures or receipts. This approach offers a more comprehensive picture of the government’s economic footprint. Despite postwar demobilization, government depredation continued to rise, indicating sustained expansion in the public sector’s relative size. Since the 1970s, however, the depredation

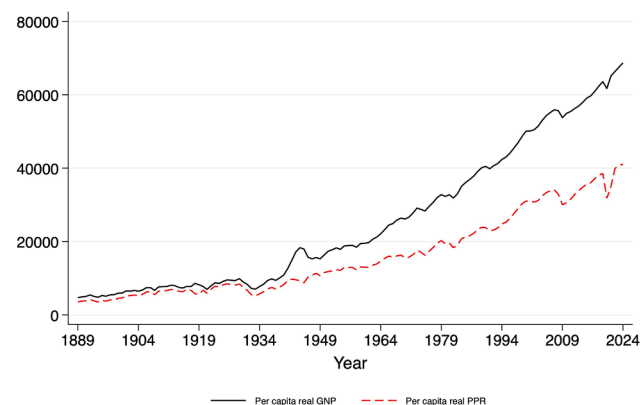


Fig. 1. Per capita real GNP and per capita real PPR in constant 2017 dollars, 1889–2024.

Table 3

Government depredation on gross national product (Percent of GNP), selected periods.

Period	Depredation (%)
1889–1928	18.38
1929–1945	28.22
1946–1973	34.06
1974–2000	40.66
2001–2024	40.73

rate has stabilized at just over 40 percent.

Building on the series shown in [Fig. 1](#), we estimate deviations of GNP and PPR from their respective long-run trends. [Fig. 2](#) plots the residuals from separate log-linear trend regressions (1889–2024) of real per capita GNP and PPR. These residuals capture how each series deviates from its long-run growth trend. Zero on the vertical axis denotes “on trend.” Positive (negative) residuals indicate that a series lies above (below) its long-run trend. It seems like the two series behave – around their respective trends – very much the same.

However, their trends differ. From 1889 to 1928, the trends are very similar (even more so for 1889 to 1916 – just prior to WWI). Thereafter, the trends differ. Thus, in addition to analyzing deviations from long-run trends, we also examine how GNP and PPR depart from their respective pre-Depression growth trajectories, following a similar approach to [Gordon \(2017\)](#). We first estimate the log-linear growth trend for GNP and PPR over the 1889–1928 period. These trends are then extended forward to 2024, allowing us to project counterfactual values for each year under the assumption that their early growth rates persisted. These projected paths serve as a benchmark against which we compare the



Fig. 2. Detrended per capita real GNP and per capita real PPR, 1889–2024.

Table 4

Ratio of actual to trend GNP and PPR based on 1889–1928 trend, selected years.

	1928	1941	1944	1950	1957	1972	1985	1995	2005	2015	2024
GNP	0.95	0.92	1.25	0.99	1.02	1.10	1.04	0.99	0.99	0.83	0.75
PPR	0.94	0.73	0.67	0.72	0.70	0.61	0.55	0.50	0.50	0.40	0.34

actual values of both GNP and PPR. Table 4 reports the ratio of actual GNP and PPR to their project 1889–1928 trend values in selected years.

The results reveal an increasingly stark divergence. While GNP meets or exceeds its pre-Depression trend during much of the mid-20th century, it falls below trend in the years following the Great Recession. This is consistent with the secular stagnation that many have discussed (Cowen, 2011; Gordon, 2017). In contrast, PPR consistently under-performs relative to this same benchmark, with the gap widening significantly over time. By 2024, PPR is only 34 % of its projected trend value, highlighting the growing share of economic activity absorbed by government and excluded from private product.

We can compare these series with other existing estimates of PPR: Rothbard (1972) for the 1929–1932 period and Batemarco (1987) for the 1947–1983 period. For both comparisons, the correlation between our series and the existing ones exceeds 0.99, indicating a very close match in levels. However, the growth rates differ somewhat. For instance, in Batemarco (1987), the estimated annual growth rate from a time trend regression on the log of GNP is 2.05%. In contrast, our corresponding estimate is 1.70% per annum. Upon further investigation, this discrepancy is attributable to the fact that Batemarco (1987) relied on a data source that has since been revised by government agencies.

4.3. Simulation of Path with a State Capacity Measures

Globally, these differences are already meaningful in the context of long-run growth estimates. More importantly, as we have argued above, the magnitude of the gap between PPR and GNP is of central importance: the nature of state capacity improvements—whether through expansion or contraction of the state—can affect PPR and GNP asymmetrically. Therefore, relying solely on aggregate output measures risks obscuring the true welfare implications of state capacity changes. This possibility is best visualized in Fig. 3 where we express the gap between GNP and PPR as a share of PPR (i.e., $(GNP - PPR)/PPR$). If the gap falls, PPR is increasing more than GNP. In such cases, as per Table 1, GNP might misstate true welfare more than PPR. In fact, given the discussion in Section 3, the decline in the gap from 1889 to the early 1920s would be particularly problematic if there was an increase in state capacity.

To illustrate the point, we employ the state capacity index developed by O'Reilly and Murphy (2022), which covers the period from 1789 to 2018. This is shown in the top panel of Fig. 4. We rescale their index such that $X \in [0, 1]$. Using Eq. (3), we estimate the position of $Z(X)$ —the legitimacy-weighted approximation between GNP and PPR—as a function of state capacity, applying different values for $\gamma \in \{0.5, 1.5, 2, 3, 4\}$. The resulting series are displayed in the bottom-left panel of Fig. 4.

As can be seen, there are noticeable differences in the level of $Z(X)$ depending on the value of γ . However, the panel is somewhat crowded, making it difficult to visually assess the relative evolution of each path. To address this, we include a bottom-right panel in Fig. 4, in which each simulated $Z(X)$ series is normalized by GNP. This yields a starker depiction of the dynamics over time: from 1889 to the early 1920s, the gap between the measures *narrows*, whereas after the 1920s and until the 1980s, it *widens*. Since the 1980s, it has shrunk again. This pattern suggests that changes in state capacity during the post-1920s period systematically alter the relationship between aggregate production and private retention, thereby inducing a potential misappreciation of the

true pace of economic growth. Consider the first phase of the changing gap—1889 to 1928. During this period, O'Reilly and Murphy (2022) document a steady increase in state capacity in the United States. However, the gap between GNP and PPR narrows, suggesting that rising state capacity was, on net, becoming less costly: it enhanced PPR more than it did GNP. This implies that the state was increasingly productive, delivering more to the private sector relative to its fiscal footprint.

This implies that using GNP during this period will *understate* true economic growth. Based on GNP, the estimated trend growth rate from 1889 to 1928 is 1.7% per annum. Depending on the value of γ , the growth rate implied by our $Z(X)$ metric ranges from 1.8% to 2.0%, while the growth rate of PPR alone is 2.1%. In the period from 1929 to 1985, state capacity continues to rise, but in forms that boost GNP more than PPR. As a result, GNP now *overstates* growth (reversing the earlier *understating* bias from 1889 to 1928). However, the absolute distance to GNP is small in that period relative to the distance to PPR. Finally, in the period from 1985 to 2018, GNP once again *understates* growth, as gains in state capacity increase PPR more than GNP.

The United States is typically classified as a high-state-capacity country. One might, as a result, expect only minimal discrepancies between PPR and GNP. After all, within the accounting identity boundaries defined by PPR (the lower bound) and GNP (the upper bound), high state capacity means a country sits closer to the upper bound and thus that GNP acts as a reliable approximation of welfare.

However, our findings indicate that even for a high-state-capacity country, the differences between are still large enough to yield significant interpretive shifts. For example, when applying our legitimacy-weighted measure $Z(X)$, the widely perceived “golden age” of growth from the late 1940s to the mid-1970s appears slower than the earlier period spanning the 1880s to the 1920s. This result is consequential because the policy frameworks associated with each era are often evaluated in light of their apparent growth performance.

In particular, the postwar period is frequently cited as a model of successful macroeconomic management, strong unions, and activist fiscal policy, whereas the earlier period is more often associated with *laissez-faire* policies and government interventions being limited to the state's regalian functions and some additional ones related to education (Goldin and Katz, 2008)³⁴ and public health (Olmstead and Rhode, 2015; Troesken, 2004, 2019). If the earlier period actually experienced faster growth when accounting for state capacity's true contribution to private welfare, then the conventional narrative—linking mid-20th-century policy to superior growth outcomes—may require substantial revision. These findings underscore the importance of measurement legitimacy in historical interpretation and in policy evaluation.

This discussion also opens an interesting possibility for future

³⁴ It is worth noting that accounts of state-provided education often overlook the supporting role played by private parochial schools in expanding access to education for Catholics—who had lower literacy rates than the general population—and in fostering competitive pressure that improved the performance of public schools (Ralph and Robinson, 1980; Gross, 2014, 2017). This possibility, though not yet formally examined, deserves attention and aligns with the claims advanced here. While the state increased its financial commitment to education, the effectiveness of that investment was likely enhanced by competition. In practice, this represents a form of state capacity improvement, albeit with ambiguous implications for interpreting the shift in measurement legitimacy.

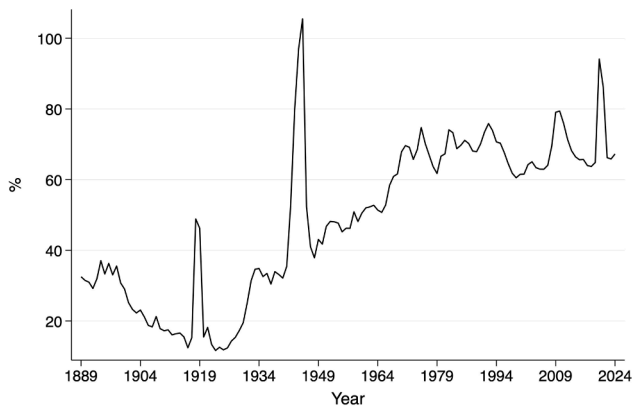


Fig. 3. Gap between GNP and PPR as a share of PPR, 1889 to 2024.

research into state capacity. Not all government spending contributes equally in terms of value-to-cost ratios. Constructing an alternative PPR measure might thus be useful if it is designed to exclude certain government sectors. For instance, a sector like education—where output is particularly difficult to value—could be excluded from the derivation of GDP, potentially yielding a lower GDP than the full measure. Alternating the excluded sectors could open the door to testing how changes in state capacity across government sectors affect the direction and magnitude of the measurement legitimacy shift.

Take, for example, policing. If the government spends less on policing because it improves the productivity of the services (e.g., less police corruption or better monitoring of police forces), PPR should increase because better policing improves productivity in the private economy, while lower taxes also increase the size of the private economy. At the same time, GDP falls (because of lower expenditures). This implies that the measurement legitimacy should move towards PPR (i.e.,

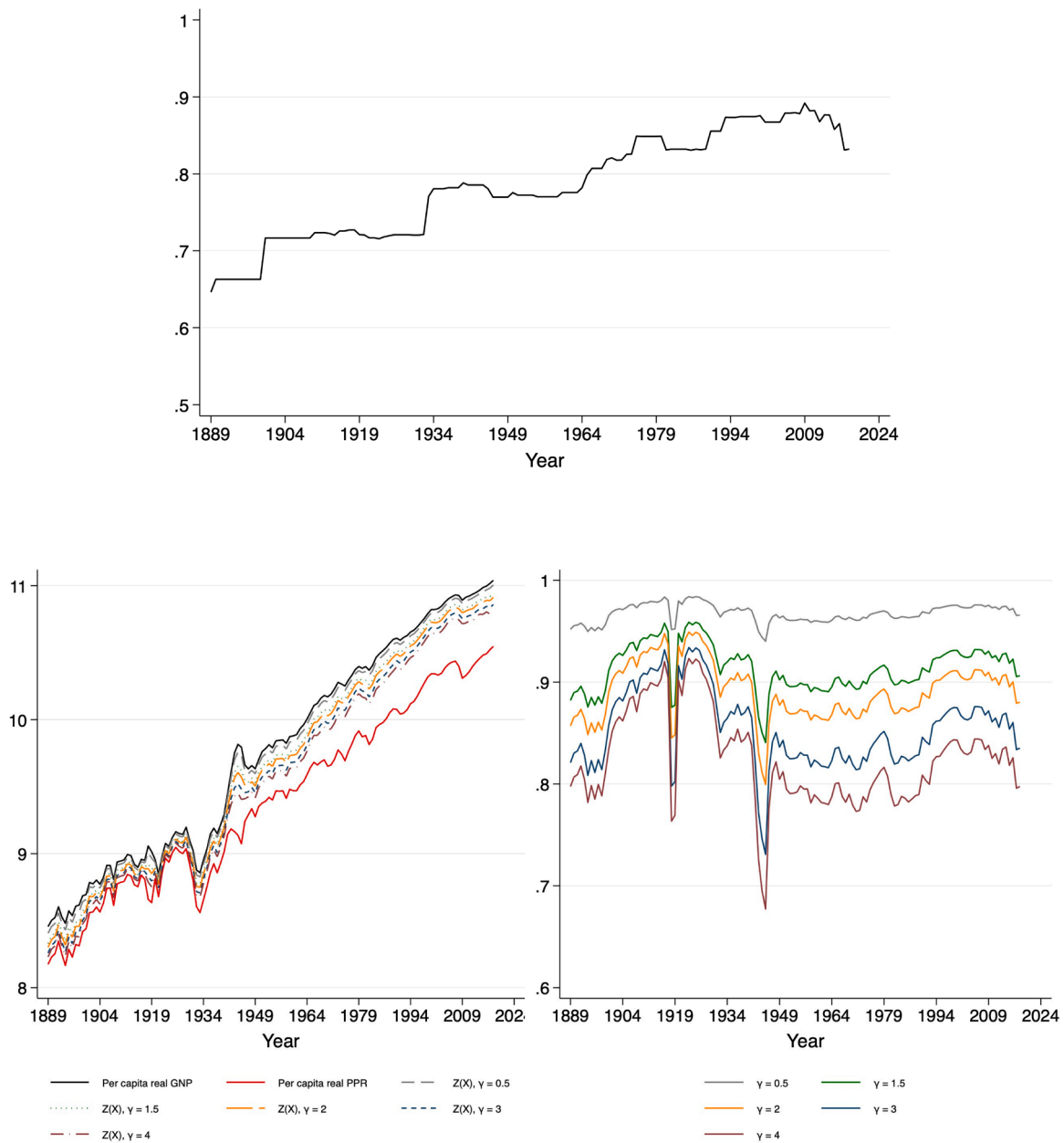


Fig. 4. O'Reilly and Murphy (2022) state capacity index for the U.S., 1889-2024 (top panel). Natural logarithm of per capita real GNP, per capita real PPR, and true welfare measures ($Z(X)$) at different values of γ indicated in the legend (bottom-left panel). $Z(X)$ calculated using the state capacity index value in each year. Ratio of GNP to value of $Z(X)$ under different assumptions of γ (bottom-right panel).

Table 5

Annualized growth rates by measure and period.

Measure	1889–1928	1929–1985	1985–2018
PPR	2.10%	2.19%	1.67%
GNP	1.68%	2.69%	1.62%
$\gamma = 0.5$	1.76%	2.68%	1.64%
$\gamma = 1.5$	1.89%	2.66%	1.67%
$\gamma = 2$	1.94%	2.64%	1.68%
$\gamma = 3$	2.00%	2.61%	1.71%
$\gamma = 4$	2.05%	2.58%	1.73%

increasing the measure's legitimacy) because we excluded one of the government sectors with genuine productivity improvements (as testified by either rising or stable output for falling costs).³⁵ In contrast, if the government spends more on policing (e.g. more officers, better equipment) in ways that generate the same improvement in PPR and increases police output, then the reverse is true—the legitimacy shift from that state capacity improvement is towards GDP. Indeed, some of the increase in GDP is from the productivity effect on PPR to which we must add the genuine value of the extra output provided by the state. For historical examples such as the ones underlying the discussion around Table 5, this would open the door to assess which sectors of government activity were most important in explaining American economic growth.

5. Conclusion

Our aim in this article was to convince economists to consider the concept of Private Product Remaining (PPR) as useful one. Developed by Murray Rothbard (1972, 2004), it measures national income that subtracts all government-generated income and the maximum of either government spending or revenues from GDP. It treats the state as a purely extractive actor whose expenditures do not contribute to economic output.

We say "useful" because PPR acts as a consistent lower boundary in attempting to value the state's contribution to economic activity. It is a lower boundary that is as consistent, expedient, and efficacious as GDP (or GNP), which serves as an upper boundary that counts all government spending as adding value. Both boundaries depict different endpoints of the state's nature in relation to its normative implications. PPR depicts a purely predatory state, where all its spending adds nothing to the productivity of the economy. GDP depicts a state led by a benevolent social planner who truly maximizes social welfare. Both extremes, obviously, do not exist in the real world. However, by knowing the boundaries, we know within which space the truth dwells.

We proposed a formalization that showed how to sniff out the correct valuation of the government's contribution to the economy. The linkage was through the concept of state capacity. Capable and constrained states can productively increase the economy's size either by providing more highly valuable outputs and thus spending more, or by cutting spending for the same basket of outputs to the public. Both paths yield a true shift in well-being, but they also induce a shift in the legitimacy (what we called the measurement legitimacy shift) of either PPR or GDP.

This, we argued, was rich in implications for the state capacity literature. Because all spending by government is considered productive, sole reliance on GDP as measurement will often misstate the effect of state capacity changes on overall welfare due to the measurement legitimacy shift. We showed that this could noticeably change – even in high state-capacity nations such as the United States – historical interpretations and narratives around development and growth.

We also think that our paper should invite major empirical revisions of key studies examining the effects of state capacity changes on development. Consider, for example, a fictitious country undergoing a reform to improve state capacity. Suppose this country is analyzed using synthetic control methods. The first step would be to determine whether the reform led to lower government spending (implying a measurement legitimacy shift toward PPR) or higher spending (implying a shift toward GDP). Second, the researcher must choose the appropriate outcome variable: PPR or GDP? If both measures show that the treated country outperforms its synthetic control, then we can be confident that the reform was effective. However, if the two measures diverge, this raises the possibility that the observed effects are driven by measurement issues. Researchers would then need to reflect on the nature of the state capacity improvement to decide which metric is more appropriate. This kind of "variable shifting" could be applied systematically across studies of state capacity and the broader relationship between government and economic development, offering a powerful form of validity check.³⁶

Overall, our work should be inviting to other scholars: it says that there is a systemic bias in a vast and growing literature on how building up states (alternatively called statecraft, state building, state capacity, state quality, constrained state capacity, productive state) can be tied to development. Other uses of PPR may be seen by other researchers—uses we have not yet identified—but we believe the one that we have identified is sufficiently important in and of itself to invite others to follow us.

CRedit authorship contribution statement

Vincent Geloso: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Chandler S. Reilly:** Writing – original draft, Writing – review & editing, Validation, Visualization, Software, Supervision, Project administration, Resources, Investigation, Methodology, Formal analysis, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

³⁵ For example, if applied to Table 5, this example would show an increase in state capacity, faster GNP growth (because we excluded a source of falling expenditures), and a smaller distance between the PPR growth rate and the Z(X) rows (with the different γ values).

³⁶ Another example would be the case of the believed quadratic relationship between government size and economic growth (Fölster and Henrekson, 2001; Di Matteo and Summerfield, 2020). Our argument implies that using GDP biases *against* finding an inverted U-curve, since higher government spending is recorded as higher GDP—even when the value-to-cost ratio of public services is low or deteriorating. The right-hand side of the data (i.e., countries with larger governments) will report measured growth that is artificially inflated (if GDP is used). This means that what should be the downward-sloping part of the curve (declining growth with excessive government) might appear flatter or even rising. Thus, the true inverted-U is obscured. Correcting for this – by shifting to PPR – might reveal a stronger and clearer inverted-U pattern, as the right-hand side would no longer be buoyed by mismeasured output.

Appendix. Constructing PPR Using GDP

Following the same procedure as described in the main text, we also construct an estimate of Private Product Remaining (PPR) that begins with GDP, rather than GNP. The only difference in data sources is that the GDP series and implicit price deflator are sourced from [Johnston and Williamson \(2025\)](#). We use the same data on government income, expenditures, and receipts to make the necessary adjustments. Rather than presenting an entire additional section presenting these results, we show the full series in [Fig. 5](#) and the detrended series in [Fig. 6](#).

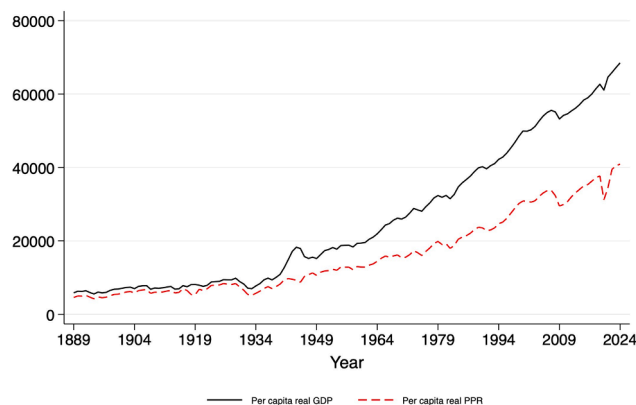


Fig. 5. Per capita real GDP and per capita real PPR in constant 2017 dollars, 1889-2024.



Fig. 6. Detrended per capita real GDP and per capita real PPR, 1889-2024.

References

- ab Iorwerth, A., 2006. How to measure government productivity: A review article on 'measurement of government output and productivity for the national accounts' (the atkinson report). *International Productivity Monitor* 13, 57–74.
- Acemoglu, D., García-Jimeno, C., Robinson, J.A., 2015. State capacity and economic development: A network approach. *American Economic Review* 105 (8), 2364–2409.
- Acemoglu, D., Moscona, J., Robinson, J.A., 2016. State capacity and American technology: evidence from the nineteenth century. *American Economic Review* 106 (5), 61–67.
- Aitken, A., 2019. Measuring welfare beyond GDP. *Natl. Inst. Econ. Rev.* 249, R3–R16.
- Anderson, T., Hill, P., 1989. *The Birth of a Transfer Society*. Hoover Institution on War, Revolution and Peace Stanford, Calif.: Hoover Institution publication. University Press of America.
- Aneja, A., Xu, G., 2022. Strengthening state capacity: Postal reform and innovation during the gilded age. Technical report. National Bureau of Economic Research.
- Atkinson, A.B., 2005. *Atkinson Review: Final Report—Measurement of Government Output and Productivity for the National Accounts*. Palgrave Macmillan, Houndmills, Basingstoke, Hampshire. Published with the permission of the Controller of Her Majesty's Stationery Office (HMSO).
- Avis, E., Ferraz, C., Finan, F., 2018. Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *Journal of Political Economy* 126 (5), 1912–1964.
- Bardhan, P., 2016. State and development: The need for a reappraisal of the current literature. *J. Econ. Lit.* 54, 862–892.
- Bateman, R., 1987. GNP, PPR, and the Standard of Living. *Review of Austrian Economics* 1 (1), 181–186.
- Bateman, R., 2023. Externalities and the state. *Quarterly Journal of Austrian Economics* 25 (4), 147–168.
- Benson, B.L., 1990. *The enterprise of law: Justice without the state*. Pacific Research Institute for Public Policy San Francisco.
- Besley, T., Persson, T., 2008. Wars and state capacity. *J. Eur. Econ. Assoc.* 6 (2-3), 522–530.
- Besley, T., Persson, T., 2009. The origins of state capacity: Property rights, taxation, and politics. *American Economic Review* 99 (4), 1218–1244.
- Besley, T., Persson, T., 2010. State capacity, conflict, and development. *Econometrica* 78 (1), 1–34.
- Boulding, K.E., 1993. *The structure of a modern economy: the United States, 1929-89*. Macmillan.
- Bournay, J., 2007. On the treatment of taxes and government in the national accounts. *Review of Income and Wealth* 53 (4), 735–746.
- Buchanan, J.M., 1975. *The limits of liberty: Between anarchy and Leviathan*. University of Chicago Press.
- Cingolani, L., 2013. *The state of state capacity: A review of concepts, evidence, and measures*. UNU-MERIT Working Papers, p. 53.
- Cowen, T. (2011). *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better*. Penguin.
- Coyle, D., 2015. *GDP: a brief but affectionate history-revised and expanded edition*. Princeton University Press.
- De Jasay, A., 1998. *The State*. Collected Papers of Anthony de Jasay. Liberty Fund.
- Di Matteo, L., Summerfield, F., 2020. The shifting scully curve: international evidence from 1871 to 2016. *Appl. Econ.* 52 (39), 4263–4283.
- Dincecco, M., 2015. The rise of effective states in Europe. *J. Econ. Hist.* 75 (3), 901–918.
- Dincecco, M., 2017. *State Capacity and Economic Development*. Cambridge University Press.
- Dincecco, M., Katz, G., 2014. State capacity and long-run economic performance. *Economic Journal* 126 (590), 189–218.
- Ellickson, R.C., 1994. *Order without law*. Harvard University Press.
- Evans, A.J., 2018. *Getting the Measure of Money: A Critical Assessment of UK Monetary Indicators*. Institute of Economic Affairs, London.

- Feenstra, R.C., Xu, M., Antoniadis, A., 2020. What is the price of tea in china? goods prices and availability in chinese cities. *The Economic Journal* 130 (632), 2438–2467.
- Fogel, R.W., Fogel, E.M., Guglielmo, M., Grotte, N., 2013. *Political arithmetic: Simon Kuznets and the empirical tradition in economics*. University of Chicago Press.
- Fölster, S., Henrekson, M., 2001. Growth effects of government expenditure and taxation in rich countries. *Eur. Econ. Rev.* 45 (8), 1501–1520.
- Forste, F., Buchanan, J.M., 1961. The evaluation of public services. *Journal of Political Economy* 69 (2), 107–121.
- Geloso, V., Makovi, M., 2022. State capacity and the post office: Evidence from nineteenth century quebec. *Journal of Government and Economics* 5, 100035.
- Geloso, V., Reilly, C.S., 2025. A defense-adjusted national accounting of the us economy and its implications. *The Review of Austrian Economics* 1791–2023 pages 1–32.
- Geloso, V., Salter, A.W., 2020. State Capacity and Economic Development: Causal Mechanism or Correlative Filter? *Journal of Economic Behavior and Organization* 170, 372–385.
- Goel, R.K., Saunoris, J.W., Schneider, F., 2019. Growth in the shadows: Effect of the shadow economy on us economic growth over more than a century. *Contemp. Econ. Policy*. 37 (1), 50–67.
- Goldin, C.D., Katz, L.F., 2008. *The race between education and technology*. Harvard University Press.
- Gordon, R., 2017. *The rise and fall of American growth: The US standard of living since the civil war*. Princeton university press.
- Griliches, Z., 1998. Productivity, r&d, and the data constraint. *R&D and Productivity: The Econometric Evidence*. University of Chicago Press, pp. 347–374 pages.
- Gross, R.N., 2014. Public regulation and the origins of modern school-choice policies in the progressive era. *Journal of Policy History* 26 (4), 509–533.
- Gross, R.N., 2017. *Public vs. private: The early history of school choice in America*. Oxford University Press.
- Hanlon, W.W., 2024. *The laissez-faire experiment: Why Britain embraced and then abandoned small government, 1800–1914*. Princeton University Press.
- Hasnas, J., 2024. *Common Law Liberalism: A New Theory of the Libertarian Society*. Oxford University Press.
- Hicks, J.R., 1940. The valuation of the social income. *Economica* 7 (26), 105–124.
- Hicks, J.R., Hicks, U.K., 1939. Public finance in the national income. *Rev. Econ. Stud.* 6 (2), 147–155.
- Higgs, R., 1992. Wartime Prosperity? A Reassessment of the US Economy in the 1940s. *J. Econ. Hist.* 52 (1), 41–60.
- Holcombe, R.G., 2004. National income accounting and public policy. *Review of Austrian Economics* 17, 387–405.
- Johnson, N.D., Koyama, M., 2017. States and economic growth: Capacity and constraints. *Explor. Econ. Hist.* 64, 1–20.
- Johnston, L., Williamson, S.H., 2025. *What Was the U.S. GDP Then?* Retrieved from <https://www.measuringworth.com/datasets/usgdp/>.
- Kalecki, M., Landau, L., 1934. *Szacunek Dochodu Społecznego, w.r. 1929 (Estimate of National Income for 1929)*, 1. Institute of Economic Research (Instytut Badania Konjunktury Gospodarczych i Cen), Warsaw.
- Knutsen, C.H., 2013. Democracy, state capacity, and economic growth. *World Dev.* 43, 1–18.
- Koyama, M., 2016. The long transition from a natural state to a liberal economic order. *Int. Rev. Law Econ.* 47, 29–39.
- Kuznets, S., 1945. *National Product in Wartime. General series*. National Bureau of Economic Research, Incorporated.
- Landau, L., 1934. *Dochody z Pracy Najemnej w.r. 1929 (Labor Income in 1929)*, 2. Institute of Economic Research, Warsaw.
- Landefeld, J.S., Fraumeni, B.M., Vojtech, C.M., 2009. Accounting for household production: A prototype satellite account using the American Time Use Survey. *Review of Income and Wealth* 55 (2), 205–225.
- Leeson, P.T., 2014. *Anarchy unbound: Why self-governance works better than you think*. Cambridge University Press.
- Lemke, J.S., 2016. Interjurisdictional competition and the married women's property acts. *Public Choice* 166 (3), 291–313.
- Lequiller, F.I., Blades, D.W., 2014. *Understanding national accounts: Second Edition, Revised and Expanded*. OECD publishing.
- Li, D.D., Maskin, E.S., 2021. Government and economics: An emerging field of study. *Journal of Government and Economics* 1, 100005.
- Migué, J.L., Bélanger, G., 1974. Toward a general theory of managerial discretion. *Public Choice* 17 (1), 27–47.
- Mueller, D.C., 1988. Anarchy, the market, and the state. *South. Econ. J.* 54 (4), 821–830.
- Niskanen, W. (1971). *Bureaucracy and Representative Government*. Aldine.
- Nomo Beyala, B.C., 2025. Do fiscal rules enhance states' fiscal capacity? *Kyklos*.
- Nordhaus, W., 2002. Productivity growth and the new economy. *Brookings Pap. Econ. Act.* 2002 (2), 211–244.
- Nordhaus, W., Tobin, J., 1972. Is growth obsolete? *Economic Research: Retrospect and Prospect*, Volume 5, Economic Growth. National Bureau of Economic Research, pp. 1–80 pages.
- North, D.C., Wallis, J.J., Weingast, B.R., 2009. *Violence and social orders: A conceptual framework for interpreting recorded human history*. Cambridge University Press.
- Olken, B.A., 2007. Monitoring corruption: evidence from a field experiment in indonesia. *Journal of political Economy* 115 (2), 200–249.
- Olmstead, A.L., Rhode, P.W., 2015. *Arresting contagion: Science, policy, and conflicts over animal disease control*. Harvard University Press.
- Olson, M., 1993. Dictatorship, democracy, and development. *American political science review* 87 (3), 567–576.
- O'Reilly, C., Murphy, R.H., 2022. An index measuring state capacity, 1789–2018. *Economica* 89 (355), 713–745.
- Piano, E.E., 2019. State capacity and public choice: a critical survey. *Public Choice* 178 (1), 289–309.
- Piano, E.E., Rouanet, L., 2020. Economic calculation and the organization of markets. *The Review of Austrian Economics* 33 (3), 331–348.
- Pritchett, L., 2022. National development delivers: and how! and how? *Econ. Model.* 107, 105717.
- Ralph, J.H., Robinson, R., 1980. Immigration and the expansion of schooling in the united states, 1890–1970. *Am. Sociol. Rev.* pages 943–954.
- Reinsdorf, M.B., Sheiner, L., 2024. *The Measure of Economics: Measuring Productivity in an Age of Technological Change*. University of Chicago Press.
- Rockoff, H., 1998. The united states: from ploughshares to swords. *The economics of World War II: Six great powers in international comparison* 81–121 pages.
- Rockoff, H., 2019. On the controversies behind the origins of the federal economic statistics. *Journal of Economic Perspectives* 33 (1), 147–164.
- Romero, D., 2025. Bureaucratic capacity and political favoritism in public procurement. *Comp. Polit. Stud.* 58 (6), 1067–1100.
- Rothbard, M.N., 1972. *America's Great Depression*. Ludwig von Mises Institute.
- Rothbard, M.N., 1978. *For a new liberty: The libertarian manifesto*. Ludwig von Mises Institute.
- Rothbard, M.N., 2004. *Man, economy, and state with power and market*. Ludwig von Mises Institute.
- Ruggles, N.D., Ruggles, R., 1956. *National Income Accounts and Income Analysis*. McGraw-Hill.
- Savoia, A., Sen, K., 2015. Measurement, evolution, determinants, and consequences of state capacity: A review of recent literature. *J. Econ. Surv.* 29, 441–458.
- Soloveichik, R., 2019. Including illegal activity in the us national economic accounts. *US Department of Commerce, Bureau of Economic Analysis*.
- Spindler, Z.A., 1982. The overstated economy: Implications of positive public economics for national accounting. *Public Choice* 38 (2), 181–196.
- Stringham, E.P., 2015. *Private governance: Creating order in economic and social life*. Oxford University Press.
- Strow, B.K., Strow, C.W., 2013. Gross actual product: Why gdp fosters increased government spending and should be replaced. *Journal of Private Enterprise* 29 (Fall 2013), 53–71.
- Studenski, P., 1958. *The Income of Nations: Theory, Measurement, and Analysis: Past and Present; a Study in Applied Economics and Statistics*. New York University Press. Number vol. 1.
- Taylor, M., 1976. *Anarchy and Cooperation*. Out-of-print books on demand. Wiley.
- Taylor, M., 1982. *Community, anarchy and liberty*. Cambridge University Press.
- Troesken, W., 2004. *Water, race, and disease*. MIT Press.
- Troesken, W., 2019. *The pox of liberty: How the constitution left Americans rich, free, and prone to infection*. University of Chicago Press.
- Tullock, G., 2005. *Bureaucracy. Obra selecta*. Liberty Fund.
- Uebele, M., Gallardo-Albarrán, D., 2015. Paving the way to modernity: Prussian roads and grain market integration in Westphalia, 1821–1855. *Scandinavian Economic History Review* 63 (1), 69–92.
- U.S. Bureau of Economic Analysis, 2025. *Gross national product [gnpa]*. Retrieved from FRED. <https://fred.stlouisfed.org/series/GNPA>. May 13, 2025.
- U.S. Census Bureau, 1975. *Historical Statistics of the United States, Colonial Times to 1970*. US Department of Commerce, Bureau of the Census. Number 93.
- Vahabi, M., 2016. A positive theory of the predatory state. *Public Choice* 168 (3–4), 153–175.
- Wagman, B., Folbre, N., 1996. Household services and economic growth in the United States, 1870–1930. *Fem. Econ.* 2 (1), 43–66.
- Wallis, J.J., 2000. American government finance in the long run: 1790 to 1990. *Journal of Economic Perspectives* 14 (1), 61–82.