NBER WORKING PAPER SERIES

IS SOCIAL PROTECTION A LUXURY GOOD?

Michael Lokshin
Martin Ravallion
Iván Torre

Working Paper 30484
http://www.nber.org/papers/w30484

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2022

The authors are with the World Bank, Georgetown University and the World Bank respectively. They are grateful to Harold Alderman, Aart Kraay, Peter Lindert, Franco Peracchi, Nithin Umapathi, Frank Vella and Dominique van de Walle for comments and/or discussions. Aylén Rodriguez Ferrari and Michael Gottschalk provided excellent assistance in data collection. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w30484.ack

Is Social Protection a Luxury Good?
Michael Lokshin, Martin Ravallion, and Iván Torre
NBER Working Paper No. 30484
September 2022
JEL No. H53,I38,O15

ABSTRACT

The claim that social protection is a luxury good—with a national income elasticity exceeding unity—has as been influential. The paper tests the "luxury good hypothesis" using newly-assembled data on social protection spending across countries since 1995, treating the pandemic period separately, as it entailed a large expansion in social protection efforts. While the mean income share devoted to social protection rises with income, this is attributable to multiple confounders, including relative prices, weak governance in low-income countries and access to information-communication technologies. Controlling for these, social protection is not a luxury good. This was also true during the pandemic.

Michael Lokshin
World Bank
1818 H Street NW
Washington, D.C. 20037
mlokshin@worldbank.org

Iván Torre
World Bank
1818 H St NW
Washington, DC 20433
itorre@worldbank.org

Martin Ravallion
Department of Economics
Georgetown University
ICC 580
Washington, DC 20057
and NBER
mr1185@georgetown.edu

1. Introduction

Social protection (SP) has long been seen as an important task for governments in rich countries but much less so in poor ones.[2] Public spending on SP tends to account for a higher share of national income in richer countries, implying an overall income elasticity exceeding unity.[3] In short, SP spending appears to be a "luxury good" at the national level. We can call this the "luxury good hypothesis for social protection," or "luxury good hypothesis" for short.

Figure 1 plots the relationship between the share of GDP devoted to SP and (log) GDP per capita across countries, mostly for 2019.[4] The line of best fit is an example of what we will call the Social Protection Engel Curve (SPEC for short), defined as the mean share of national income devoted to SP, conditional on national income.[5] Consistently with the luxury good hypothesis, we see that the SPEC rises through almost the whole range of GDP (only turning down when we reach the five countries with the highest GDP per capita). We see a large difference between rich and poor countries in the mean shares of GDP devoted to SP; the share is 15% on average among the top quintile of countries ranked by GDP per capita (albeit with a large variance), as compared to around 3% for the bottom quintile.

The luxury good hypothesis for social protection has been influential in development policy discussions. Evidence such as Figure 1 has been used to argue that poor countries have higher priorities than SP. Indeed, SP is not even mentioned in the classic policy package for developing countries that emerged in the 1990s, often referred to as the "Washington Consensus," following Williamson (1990). Williamson identifies health and education as "proper objects of government expenditure," but says nothing about SP. The more explicitly poverty-

---

[2] Following World Bank usage, SP refers to public programs for "social assistance" (in the form of cash or in-kind benefits often targeted to poor families), "social insurance" (mainly unemployment benefits and pensions), and "active labor market programs" (such as job training schemes). On the history of thought on the role of SP see Ravallion (2016a, Part 1).

[3] See, for example, Kristov et al. (1992), Lindert (1994), Peracchi (1998), Cornelisse and Goudswaard (2002), Auteri and Costantini (2004), Shelton (2007), Brückner et al (2012), and Clemente et al. (2012). The finding that public spending as a whole tends to rise as a share of GDP as GDP rises is sometimes referred to as Wagner's Law.

[4] Figure 1 uses the most recent pre-pandemic data on total SP spending for each country. We consider the pandemic period separately. We also consider the effect of dropping spending on pensions; the pattern in Figure 1 is very similar in this case (Addendum).

[5] As estimated by a nonparametric regression function in Figure 1. Our use of the term "Engel curve" borrows from the more familiar usage in the context of modeling consumer demands. We are not the first to use the term in the context of public spending, which appears to be Bird (1971). Figure 1 and other SP Engel curves in this paper are nonlinear (and non-parametric) generalizations of the longstanding Working-Leser specification for such Engel curves, for which the share of total income (or total spending) is taken to be linear in log total income (Working 1943; Leser 1963).

focused policy discussions of the 1990s gave more attention to SP, although it was seen as a short-term palliative to help mostly middle-income countries address downside risks and help specific disadvantaged groups (such as the disabled) who could not be expected to benefit directly from economic growth.[6] This policy prioritization regarding SP often came alongside a view (not always explicit) that governmental efforts at reducing inequality (including SP but not confined to SP) in poor countries should wait until the countries are not so poor. Poverty reduction was perceived as the near-term goal, and economic growth in a market economy (supported by sound public investments in health and education) was seen as the primary instrument.

There are signs that this view of SP has been changing. Over the last 20 years or so we have seen greater emphasis on the development role of SP in low- and middle-income countries, as an "equal partner" with more traditional development instruments. The volume of development assistance for SP by the World Bank and most other international development agencies (both multilateral and unilateral) rose appreciably alongside the change in thinking. As a prior Director in charge of the World Bank's social protection efforts, Xiaoqing Yu (2016), put it, "Social protection is no longer considered to be a luxury." Nonetheless, as Figure 1 illustrates, the pattern of a rising share of GDP devoted to SP as GDP rises is still evident in recent data. (Possibly the SPEC would be even steeper if we looked at data for, say, 25 years ago; we shall return to this point.)

However, the luxury good hypothesis is not the only explanation for the pattern seen in Figure 1. This paper tests the hypothesis against seemingly plausible alternatives. The question is why we see higher SP shares among countries with higher GDP—a rising cross-country SPEC, as in Figure 1. Is it coming from how higher national income influences demand for SP, or does it stem from other (omitted) characteristics that are jointly correlated with national income and the SP share? The answer matters to how one interprets the SPEC. The "luxury good" interpretation has been taken to imply that the rising SPEC stems from preferences—people in poor countries care more about other things than SP. The alternative view is that they care just as much (or possibly more) about SP but weak institutions, deficient infrastructure and other

[6] For example, the World Bank's (1990) influential World Development Report, *Poverty*, identified a role for SP, but this was (explicitly) secondary to the report's two-part strategy for sustained poverty reduction by combining policies to encourage labor-intensive economic growth in a market economy with investments in health and education.

conditions typical of poor countries, make it harder to convert their notional SP demands into public spending outcomes—in short, that poorer countries face higher costs in effectively implementing this type of public spending. Prices of competing human development services (health and education) may also play a role, to the extent that these services are relatively expensive in richer countries, encouraging substitution toward SP transfers within public budgets.

Testing the luxury good hypothesis can be readily recognized as a classic issue in causal inference. The idea of omitted variable bias (OVB) provides a natural framework for structuring our effort. We formulate a series of alternative hypotheses to the luxury good hypothesis. These suggest specific country characteristics that might be expected to influence the SPEC while also being correlated with national income. (We strive to avoid a "kitchen sink" approach to controls, but instead to carefully formulate hypotheses based on the literature or our reasoning.) We interpret these characteristics as reflecting differences in the implicit prices faced in providing SP; for example, that the implicit price of SP is higher when government effectiveness or technology access in general is lower. We do not claim that our list of potential confounders is exhaustive, but (as we will see) it is enough to be confident about the answer to the main question posed by this paper.

The specific hypotheses we consider relate to relative prices for human development services, political economy (governmental accountability, the capacity for public service provision more broadly, and redistributive policy making), the distributional impacts of economic growth, access to appropriate technologies, population ageing, and the selection processes for observed public spending data. We argue that some of the hypotheses we consider can already be dismissed from what we know, based on the literature. Among the rest, each hypothesis points to one or more potential confounders in estimating the SPEC, which we use in testing for OVB. We also point to some likely theoretical ambiguities in how the confounders can influence SP spending; for example, having a more capable government can be expected to reduce the unit cost of attaining a given level of social protection (thus reducing SP spending), but also increase the desired level of protection (increasing spending).

The paper uses two new data sets on public spending on SP. The first spans 142 countries over 1995-2020, combined with data on identified covariates. The second data set focuses on the pandemic period 2020-21 and covers SP programs under the designated pandemic stimulus budget of each of 154 countries.

Our motivation in using this second data set is that, prior to the pandemic, the leadership of low- and middle-income countries may have wanted to see more resources devoted to SP, but faced "stickiness" in budget allocations and difficulties in domestic resource mobilization (see, for example, Alesina and Passarelli 2019). There may well have been a genuine change in thinking about SP among policy makers, but it just takes a long time for this to be evident in the SPECs. A major shock can potentially break the hysteresis and so reveal the new priorities. Such a shock can also be expected to systematically influence the composition of SP spending; for example, the pandemic encouraged greater emphasis on job protection/retention schemes (Gentilini et al. 2021; Demirgüç-Kunt et al. 2022). We ask if this opportunity was taken up by poor countries, such that SP ceased to be a luxury good during the pandemic.

What do we find? We first show that the positive income effect on SP shares evident in Figure 1 remains when we introduce country fixed effects; SP is still a luxury good when we focus on the inter-temporal variances, although the slope of the SPEC is attenuated considerably. Next, we argue that the tendency for the SP share to rise with national income is mainly attributable to OVB stemming from multiple, time-varying, confounders. When we also include country fixed effects, the SPEC that we get after controlling for the confounders turns out to be negatively sloped with respect to GDP per capita, contrary to the luxury good interpretation.

In studying the policy responses to the pandemic, we do not find that the opportunities created for implementing a radically new policy regime (with greater emphasis on SP in poor countries) were taken up in general. Relatively low public spending on SP among poorer countries during the pandemic appears to stem mainly from weak government effectiveness in public service delivery, and younger populations, rather than low income *per se*, again rejecting the luxury good hypothesis.

The following section outlines the series of alternative hypotheses. Section 3 describes the data we have assembled. Section 4 presents our results on testing whether one or more of our alternative hypotheses can explain why we see higher SP spending as a share of GDP in richer countries. In Section 5, we turn to the SP responses to the pandemic. Section 6 concludes.

2. Alternative hypotheses

There is more than one way to interpret the finding that poorer countries tend to devote lower shares of their national income to SP (as seen in Figure 1). The "luxury good" interpretation assumes that public spending choices are reasonably responsive to the preferences of citizens. Then Figure 1 is seen to stem from a preferred policy hierarchy that simply reflects the preferences of citizens, most of whom (it is inferred) consider SP to be a luxury good. The take-away for policy is that the relatively low level of SP spending in poor countries should not be of much concern. If citizens of poor countries would rather see public spending going elsewhere, why should we care?

However, there are a number of alternative explanations for the rising SPEC, with different implications for development policy. These can be thought of as differences in the costs incurred by governments in supplying SP, as reflected in omitted country characteristics in a static cross-country comparison, such as Figure 1. This can be addressed in part by introducing country fixed effects into a SPEC estimated on a longitudinal data set. However, there are also concerns about omitted time-varying characteristics correlated with GDP. We can motivate thinking about those omitted characteristics in the form of a series of hypotheses, as follows. We note if the existing literature already includes what appears to be an adequate basis for rejecting the hypothesis and proceeding no further. For the rest, we proceed to further empirical testing. Our hypotheses can be grouped under four headings: relative prices, governance, distribution, technology and demographics.

The first hypothesis starts from the long-standing observation that Engel curves can shift with relative prices. While SP does not have its own "relative price" in any obvious sense, there are explicit prices for substitutes for SP within the gamut of public spending on human development. For 2017, the price data (described later) indicate that the price of health services relative to food and non-alcoholic beverages has a correlation coefficient of 0.61 with (log) GDP per capita; for education services the correlation coefficient is 0.62 (The Addendum provides the scatter plots, Figure A1). Given that health and educational services are more costly in richer countries we expect to see substitution in favor of SP within public budgets. This leads to:

Prices: The relative prices hypothesis. *Higher relative prices of competing human development services in richer countries lead their governments to switch to SP transfers.*

Our next set of hypotheses relates to governance and political economy, starting with:

Governance 1: The accountability hypothesis. *SP is equally important for people from poor and rich countries, but the governments of poorer countries tend to be less responsive to their citizens' demands.* An autocratic dictator likely faces fewer incentives to respond to citizens' demands (at least once the local elite is protected) than a government in a robust democracy. Even among democracies, some have better institutions (such as through greater freedom for the media) to help assure accountability.

The literature has provided mixed evidence related to Governance 1. Lindert (1994) argued that the rise of democracy, alongside economic development, was an important factor in the increase in SP spending in today's rich world when it was not nearly so rich. However, Mulligan et al. (2004) find little sign that democracies tend to pursue different social policies, once one controls for other economic and demographic differences. Dincecco (2009) argues that more absolutist (and presumably less accountable) political regimes in pre-World War 1 Europe tended to raise less revenue generally, which would have constrained SP spending. We shall explore this hypothesis further, as well as its companion:

Governance 2: The government effectiveness hypothesis. *Poorer countries have less capable governments for delivering public spending, whether on SP or something else.* The literature on institutions and development has often pointed to ways in which being a richer country can promote better institutions and polices.[7] The cost of delivering a unit of social protection is presumably lower in countries with better institutions. That can explain Ravallion's (2017) finding that the coverage of social protection is positively correlated with national income. Government effectiveness is likely to be correlated with GDP, but it also has its own independent variation, in that there are poor countries with relatively effective governments, and some rich countries for which one could not reasonably say that. This is illustrated by Figure 2, which plots the "government effectiveness (GE) index" from the World Bank's *Worldwide Governance Indicators* (WGI) against GDP per capita. The GE index is highly correlated with GDP per capita. However, there are low-income and middle-income countries with relatively high GE values and high-income countries with low values. (For example, Saudi Arabia's GE index is about the same as Rwanda's.)

[7] Thus, this strand of the literature gives much attention to the likely endogeneity of institutions in explaining differences in rates of economic growth; see, for example, Acemoglu et al. (2005).

The last of our three hypotheses under the governance heading concerns local resource mobilization:

Governance 3: The domestic resource mobilization hypothesis. *Poorer countries are more constrained in raising revenues needed for financing SP spending.*[8] The key assumption here it that the easier it is for the government to mobilize resources domestically through taxation, the easier it will be to finance transfers as SP. By contrast, most external funding sources (such as commercial and development Banks) are more demanding that the monies go toward things that are traditionally seen as investments, with a pecuniary return that would help assure loan re-payment. (When available, development grants appear more amenable to spending on SP.)

Governance 2 and 3 clearly overlap. The public services that a government is able to provide depend on its ability to finance them through domestic taxation, though also borrowing or grants, including development assistance. As noted, the accountability of governments might also be expected to influence resource mobilization. One might subsume Governance 3 within Governance 2, but we think it is likely to be of sufficient importance to warrant its own reckoning as a covariate of SP spending.

Next, we consider three hypotheses related to the distribution of income.

Distribution 1: The median-voter hypothesis. *More skewed income distributions in rich countries encourage spending on SP as a form of redistribution.* The premise here is that more (positively) skewed income distributions encourage more redistributive public spending as formalized in the famous median-voter theorem (Meltzer and Richard, 1981). Social protection is, to some degree, a redistributive intervention. Skewness in the income or wealth distribution can then be expected to matter to the political economy of SP spending. Additionally, if poorer countries tend to have less skewed income distributions, then one might expect to find that SP spending is a lower priority in those countries.

There are reasons to question this hypothesis. One reason is that some (mostly developing) countries are not democracies, so the voting mechanism postulated by the median-

[8] This can be seen as an instance of the more general argument that the capacity for domestic taxation is important for development outcomes, as in Besley and Persson (2011).

voter theorem is absent. However, it is not clear that this would fully neutralize the relevance of income distribution; even dictators cannot entirely ignore their citizens' wishes.[9]

A more compelling reason to question this hypothesis is that the evidence suggests that poorer countries tend to have more unequal distributions.[10] The (small) literature on the role of distribution in influencing public spending has almost solely used the Gini index (as in, for example, Shelton 2007), rather than skewness, but the two are likely to be highly correlated. Empirically, we find that skewness is actually a more common feature of income distributions in low-income countries than in high-income ones; we find that the correlation coefficient between (log) GDP per capita and the ratio of the mean to the median incomes is -0.405.[11] Thus, if anything, the political economy of redistribution (or at least the median-voter version) implies more redistributive effort in poorer countries. So, we appear to know enough already to reject this hypothesis as an explanation of the rising SPEC, and we will not test it further.

Distribution 2: The unequal growth hypothesis. *Economic growth in market economies tends to be inequality increasing. This leads to political demands for redistributive SP policies.* The idea that higher inequality creates demands for redistribution is familiar from the political economy literature (including, for example, Alesina and Rodrik 1994). In the present context, the concern is that the positive income effect on the SPEC is actually picking up the effect of economic growth on inequality, leading to higher SP spending.

However, in light of accumulated evidence from household survey data, the basic premise of this hypothesis can be questioned. Empirically, we see a far more mixed picture, with inequality decreasing roughly as often as it increases in growing economies.[12] These findings have been mainly from research on developing countries. However, a similar result is found when one includes high-income countries; in the global data set we use below, the correlation coefficient between the first difference of the log Gini index and the first difference of log GDP per capita is -0.09.[13] So, from what we know already, this hypothesis can also be dismissed.

---

[9] Han (2021) finds that in authoritarian regimes that hold elections, redistributive policies are implemented in the runup to elections even when election results are predetermined.
[10] See, for example, World Bank (2006) and Milanovic (2016).
[11] We measure the ratio of the mean to the median, with both measured from the household surveys, as processed in the World Bank's *PovcalNet* site. While the ratio of the mean to the median is not strictly skewness (the third moment of the distribution), it is an intuitively appealing indicator. The ratio exceeds unity in all countries as one would expect.
[12] Ravallion (2016a, Chapter 8) reviews the evidence on this point. Note that we refer here to relative inequality; absolute inequality tends to rise with growth; on this distinction and the evidence see Ravallion (2021).
[13] For the levels regression with country fixed effects the partial correlation coefficient is -0.03.

One goal of SP is to assure some minimum standard of living in the society, as typically indicated by a poverty line. Naturally, national SP efforts depend on prevailing national poverty lines rather than international absolute lines, which aim to have common real value across countries (such as the World Bank's $1.90 a day line). This motivates our next hypothesis:

Distribution 3: The relative poverty hypothesis. *Higher national poverty lines in rich countries imply that more public SP spending is needed to reduce poverty.* In assessing the *a priori* case for this hypothesis it must first be recognized that there are two, potentially offsetting, ways in which a higher average income will alter the cost of reducing poverty using SP spending. Poverty lines are typically relative in rich countries, meaning that they rise in proportion to the mean or median income; low- and middle-income countries, by contrast, tend to have absolute or only weakly relative poverty lines, meaning that they have an elasticity to the mean that is less than unity (Ravallion et al. 2009). On the other hand, the national poverty gap index (PG) using weakly relative poverty measures is likely to be a decreasing function of mean income.[14] It is then an empirical question whether the aggregate poverty gap per capita (PG times the national poverty line) is increasing or decreasing with mean income. If the (positive) elasticity of the national poverty line to mean income exceeds (in absolute value) the negative elasticity of the national poverty gap index with respect to mean income, then the cost of targeted efforts in using SP to reduce poverty will tend to be higher in rich countries.

To test this we can draw on estimates from the literature of these two elasticities. Ravallion (2016b, Appendix) estimates the cross-country elasticity of the national poverty line to the mean to be 0.52 (robust s.e.=0.04; n=598).[15] Using the estimates of the relative poverty gap index at the country level (calibrated to predicted national poverty lines) obtained by Ravallion and Chen (2019), and also allowing for country fixed effects, we obtain an elasticity with respect to the mean income or consumption of -0.56 (s.e.=0.08; n=144). Thus, the two effects of higher average income on the national poverty gap per capita (PG times the poverty line)—one on the national poverty line and one on the PG itself—essentially cancel out, suggesting that the income elasticity of the cost of eliminating poverty is likely to be close to zero.

[14] PG is the aggregate proportionate distance below the poverty line (expressed as a proportion of the line and counted as zero for those above the line) per capita of the total population.

[15] This is based on the compilation of national poverty lines across countries and over time from Jolliffe and Prydz (2016)). Ravallion (2016b, Appendix) regressed the log of the national poverty line on the log of mean income including country fixed effects.

In short, based on the existing evidence, we do not consider the relative poverty hypothesis further, since we do not expect the cost of eliminating poverty to be correlated with mean income.

A further hypothesis relates to access to technology for supporting SP programs:

Technology: The ICT hypothesis. *The governments of poorer countries have less access to ICT, which makes SP programs more costly options for public spending.* The literature on SP has not often discussed this issue, but it is a seemingly plausible assumption that SP programs tend to be ICT-intensive, such as in creating and maintaining social registries of participants and payment and monitoring systems. There is evidence that such ICT capabilities enhance the quality and coverage of SP programs in poor countries.[16] Better communication infrastructure in a country can facilitate new options for using mobile money technologies that have proven to be helpful for cash transfer delivery in some low-income settings.[17] However, with limited access to ICT at the national level due to poor infrastructure, the cost of implementing SP programs can be expected to be higher in poorer countries, leading policy makers to substitute SP in favor of other types of public spending. We will test this hypothesis using ICT usage indicators at the national level over time.

The next hypothesis concerns population aging:

Demographics: The aging hypothesis. *Richer countries tend to have older populations, which increases the demand for SP.* The fact that richer countries tend to have more elderly populations follows almost immediately from the well-known Preston (1975) curve, whereby life expectancy at birth tends to be an increasing (concave) function of average income across countries.[18] With an older population in richer countries we can expect to see demands for extra public spending to help support the elderly, who have diminished capabilities for supporting themselves; Shelton (2007) finds cross-country evidence to support this claim. Since pensions are one component of SP spending, countries with an older population will tend to have higher SP spending, though non-pension components may also be higher. A preliminary look at the evolution of SPECs excluding pension spending (Addendum Figure A3 and Figure A4) does not suggest that pension spending explains the pattern in Figure 1. We will test the aging hypothesis.

---

[16] For examples of the use of ICT in registration, payments and other functions, see World Bank (2022c).
[17] See, for example, Jack et al. (2013).
[18] For more recent descriptions and analysis see (inter alia) Gómez and Hernández de Cos (2008) and Ritchie and Roser (2019).

A further hypothesis relates to potential sample selection bias in the data underlying Figure 1 and similar figures in the literature. We will be using panel data on SP spending across countries and over time. However, it is an unbalanced panel, in that not all countries have data in all years. The potential for a non-random pattern in the availability of data suggests another reason we see a rising SPEC. It is a plausible assumption that the latent selection process increases the observed mean share of income devoted to SP spending conditional on mean income; all that is required for such positive selection is that higher SP/GDP values are more likely to be reported. However, that is not sufficient to explain why we see a rising SPEC. For that step in the argument, we draw on the literature on estimating models with sample selection bias.[19] Following that literature, let $g(p)$ denote an additive control function for the selective reporting, which is taken to be some suitably smooth function of the probability of being selected into the sample, denoted p.[20] Then, with positive selection ($g'(p) > 0$), the key issue in the present context is the sign and size of the correlation between the control function for selection and GDP per capita.

In thinking about that correlation, it is instructive to consider a simple economic model of how selective reporting arises in this context. Consider the costs and benefits of providing public spending data. (Without loss of generality, these can be taken to be the costs and benefits as perceived by public decision makers charged with this task.) It is a plausible assumption that the marginal cost (MC) to a country's government of collecting, processing and disbursing such data rises with the amount of data made available. The marginal benefit (MB), on the other hand, can be expected to fall if anything, as diminishing returns set in to collecting and processing extra data. It is also plausible that the MC of providing data tends to be higher for low-income countries than for high-income ones. This reflects the lower levels of state capacity generally in poorer countries, stemming from weaknesses in both governance and ICT. The government chooses a level of data availability equating MC and MB, with the implication that low-income counties tend to make less data available as the MC function shifts upwards.

Then we can expect the control function for selection, $g(p)$, to be positively correlated with GDP per capita. However, if one ignores the non-random selection then the control function

---

[19] See Vella's (1998) survey.

[20] More precisely, $p(\boldsymbol{Z}) = Prob(d = 1|\boldsymbol{Z})$ is the propensity score for selection, where $d = 1$ if the country, year combination is observed while $d = 0$ otherwise, and $\boldsymbol{Z}$ is a vector of covariates for selection.

is an omitted variable, positively correlated with GDP per capita. Thus, with the addition of this economic model, we have another alternative to the luxury good hypothesis:

Selection: The selective reporting hypothesis. *Richer countries tend to have more complete SP spending data which imparts an upward bias to the estimated income slope of the SPEC.* We test this hypothesis.

Lastly, in the context of the pandemic, we consider a further hypothesis:

The COVID hypothesis: *Richer countries were impacted more by the pandemic and responded accordingly with their own SP spending.* By this view, rich countries had more of a shock to protect themselves from, reflecting higher levels of social and economic interaction that helped spread the infection (prior to vaccine availability). We can use COVID mortality data in measuring the severity of the pandemic shock.[21] These data do suggest a greater COVID impact in countries with higher GDP per capita; in our data set, the correlation coefficient between COVID deaths per capita and GDP per capita is 0.439. We test this as an extra covariate for SP spending during the pandemic.

3. Data and descriptive statistics

The main dependent variable in our analysis is the share of national income devoted to public SP expenditure. In our first set of data, the coverage is across countries and over time up to (but not including) the pandemic, while the second is for the designated SP responses to the pandemic.

3.1 Pre-pandemic data on social protection

Data on public spending at the country level were assembled for the purpose of this paper from the records kept by multiple agencies, notably Eurostat, OECD, ECLAC, IMF, and the World Bank. We have drawn on all these sources to provide as complete a data set as feasible. The Addendum provides details on the specific sources for every country and year, covering 142 countries over 1995-2020 (Table A1).

[21] We acknowledge that these data may be subject to measurement error, especially in many low-income countries. Another indicator of the severity of the pandemic is the change in GDP during the pandemic, but this is endogenous to the SP efforts in response.

There are limits to how far cross-country comparisons, with limited degrees of freedom, can be considered conclusive about the determinants of the SPEC. However, by pooling cross-country observations with time series data, we will have 2,481 observations for assessing the relative importance of these various factors. On average, we have 17.5 time-series observations per country with non-missing shares of SP expenditure. We focus mainly on total SP spending, though we note any important differences between pensions and other SP spending, since public pension schemes often appear to be serving a somewhat different role to other types of SP spending (notably the more poverty-focused category of "social assistance"). The sources allow us to distinguish pension from non-pension SP spending for more than 1,500 observations across 104 countries, although coverage is less complete for low-income countries.

The panel data are unbalanced. A saturated panel would have 3692 (=142 x 26) observations. So, we have about two-thirds (0.67=2481/3692) coverage for the basic SPEC (though this falls further when we introduce covariates). The number of country observations per year is increasing, from 75 in 1995 to a maximum of 112 in 2017. In no year do we have SP spending data for all 142 countries; 45 countries have 26 years of data, and 73 have 20 or more years. The Addendum provides a series of graphs (Figure A2) that show the missing observations are more likely to be lower-income countries. (A similar pattern is found for most other covariates discussed below.)

3.2 Descriptive statistics on SP spending

Some simple descriptive statistics based on the (pre-pandemic) panel data set are instructive. On average, we find that SP spending has risen over time as a share of GDP, and that this has been more pronounced among initially low- and middle-income countries. This can be seen in Table 1, which tabulates the SP spending share by countries classified according to their initial (1995) GDP per capita. For the lowest quintile of countries, the share of GDP going to SP increased by a factor of three between 1995 and 2015, from 1.4% to 4.2%.[22] By contrast, the share has been quite stable over time for the upper two quintiles.

The rise in the SP share among initially poorer countries seen in Table 1 could reflect either economic growth in those countries or a change in policy priorities, whereby the share rose

[22] Much of the rise in SP spending by initially poorer countries was coming from public outlays on pensions; the Addendum provides details in Table A2.

at a given level of GDP per capita, implying an upward shift in the SPEC at the lower end. That can be assessed using Figure 3, which augments Figure 1 to provide the nonparametric SPECs for various years. We see that, if anything, the SPECs have tended to shift downwards over time. Among low-income countries, we see little or no change in the SP share at a given level of GDP per capita except for a drop in average levels between 1995 and 2000. In the middle-upper income range, we actually see a decline after 1995. Across all levels of GDP, the 1995 SPEC is unambiguously higher than those for 2005, 2010, 2015 and 2019, while the 2019 SPEC is unambiguously lower than those for all years except 2000 and 2005. The greater emphasis on SP in development policy discussions in the new Millennium does not appear to have shifted the SPEC upwards.

This finding is not inconsistent with the recent expansion in SP programs in developing countries; it is just that these programs have typically had low coverage (on average) in low-income countries, as documented in Ravallion (2017). The rise in SP spending as a share of GDP in developing countries is coming from higher GDP rather than a change in development priorities in favor of a higher share of national income going to SP at given GDP.

3.3 Data on social protection spending during the pandemic

Our panel data set on SP spending prior to the pandemic is augmented by a database compiled by a team including two of the present authors and documented in Demirgüç-Kunt et al. (2022). This provides estimates of the expenditures incurred in social protection measures that the government implemented in response to the pandemic over 2020 and 2021 in 154 countries. The data draw on information from the Global Database on Social Protection Responses to COVID-19 (Gentilini et al. 2021), budget data from official documents (including IMF Article IV revisions and other international organizations' related documents), government websites, and news sources. The key criterion for the inclusion is that the component of SP spending must have been designated as a response to the pandemic. In some cases these were new programs while others were expansions to existing programs. To the extent feasible, we estimated the budget corresponding to the expansion of an existing program (for example, the cost of extending the eligibility period of unemployment benefits) and not the whole pre-existing program, though we cannot always be certain of this since sometimes the source was ambiguous about this difference.[23]

3.4 Data on the covariates

The available data limit our ability to test the Relative Prices Hypothesis. The International Comparison Program provides estimates from each ICP round of the prices of different consumption categories, including health services, education services and the aggregate category of food and non-alcoholic beverages. With this information, the relative price of health and education to food and non-alcoholic beverages can be easily calculated.[24] However, the ICP rounds are not annual: they are only available for years 2005, 2011 and 2017.[25] The health and education price relatives are highly correlated (r=0.92 for 2005, r=0.90 for 2011 and r=0.82 for 2017), so we show results with them entering separately as well as jointly.

In testing Governance 1, thanks to the Word Bank's WGI, we have regular measures of the responsiveness of governments to their citizens, namely their "voice and accountability" (VA) index. As defined on the WGI website, VA "… captures perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media." The index is a composite of indicators on democracy, electoral processes, accountability of public officials, rights, reliability of government budget documents, transparency in policy making, the freedoms allowed for citizens, and citizens' trust in their government. The index comes normalized to have mean zero and a standard deviation of unity in every year.

We consider two other measures of accountability. The first is the electoral democracy index (also known as the polyarchy index) produced by the Varieties of Democracy (V-DEM) project. This measures the extent to which a country has free and fair electoral competition, suffrage is extensive, political and civil society organizations operate freely, there is freedom of expression and the media is independent. The index ranges from 0 to 1, where zero indicates a regime furthest away from electoral democracy and 1 is a full electoral democracy. The second

---

[23] We checked against the IMF estimates on the total additional public spending (whether on SP or something else) implemented during the pandemic (IMF 2021); in no case did our estimate exceed the corresponding IMF number.
[24] The ICP provides information of the price level of each category indexed to the world average, which takes a value of 100. The ratio of these price levels provides an indicator of relative prices.
[25] The number of countries for which social protection and ICP data is available is 81 in 2005, 99 in 2011 and 102 in 2017.

measure is the political competition index produced by the Polity5 project (Center for Systemic Peace 2020). This categorical index measures the degree of political competition in a given country, independently of the political regime, and ranges from a value of 1 (meaning that political competition is suppressed) to a value of 10 (meaning that political competition is institutionalized and electoral). We use this measure as an additional test of the Accountability Hypothesis under the assumption that regimes where political competition is higher tend to be more accountable and responsive to citizens' demands.

In testing Governance 2, we again turn to the WGI, which also provides a measure of government effectiveness (GE) across countries. To quote the WGI website, "Government effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (World Bank 2022b). GE is a composite index of indicators of the quality, coverage, and citizen satisfaction with public goods and services (roads, public transport, electricity, health and education services, water and sanitation, bureaucratic quality). Similarly to VA, the GE index is normalized to have mean zero and a standard deviation of unity.

For the Technology Hypothesis, we draw on data from the International Telecommunications Union (ITU) that focuses on the use of ICT.[26] We focus on two technologies, namely mobile phones and the internet. For the former, we use the number of mobile phone subscriptions per capita. For the latter, we use the percentage of the population using the internet. Both these measures are only available from the year 2000 and onwards.

While we have about 2,500 observations of the expenditure on SP by country and year, the covariates used to test the hypotheses laid out in Section 2 sometimes have a shorter time period and more limited geographical coverage. The subsample for which all variables except relative prices are available includes about 1,500 observations covering 107 countries. Relative prices are only available for three years (2005, 2011 and 2017) and therefore the subsample including them (in addition to the remaining covariates) is limited to 220 observations from 95 countries.

[26] One might prefer to test this hypothesis with information on the use of ICT services by each country's government, but there is no data source that provides enough time and geographical coverage on that dimension.

4. Tests of the rival hypotheses for why we see a rising SPEC

The observations in Section 3.2 point to the role of economic growth in fostering higher SP spending in developing countries. This requires a rising SPEC, whereby shares of GDP devoted to SP tend to be lower in countries with lower GDP, as in Figure 1. But our main question is now begging: Why do we see the rising SPEC?

It is well-recognized that cross-country comparisons such as in Figure 1 can be deceptive given that country characteristics could be correlated with national income and also influence SP spending. If one has not controlled for those characteristics then GDP is endogenous in the estimation of the SPEC. We address this concern by adding controls for (time-varying) observables and using the panel structure of our data to deal with (time-invariant) unobservables, assuming that the correlation between the error term of our regression model for the SP share and the regressors is fully captured by an additive country fixed effect.[27] Under this assumption, we shall estimate the SPEC by pooling years and countries, and including country fixed effects, thus relying solely on the inter-temporal variances. We acknowledge that biases may remain due to time-varying unobservables (or measurement errors). The textbook solution is to use one or more instrumental variables (IVs), but we are skeptical of the possibility of finding valid IVs in this context.[28] Careful selection of the control variables appears to be the best option.

Section 2 pointed to a number of potentially confounding variables correlated with levels of GDP (or for which changes over time are correlated with growth rates). Table 2 summarizes the various hypotheses, including the control variables (when further testing appears to be warranted based on what we know already from the literature). We build up our tests by adding control variables to the basic SPEC in which the SP spending share depends solely on GDP per capita. We initially add each candidate variable one at a time, to see how far each one can take us on its own in explaining the income effect on the SP share. The expectation is that the potential confounder will reduce the coefficient on (log) GDP per capita and possibly eliminate its effect.

[27] We expect that any reverse causality, whereby GDP responds (positively or negatively) to SP spending, is a longer-term effect, such that the short-term changes in GDP (after allowing for time-varying controls and country fixed effect in our panel data set) can be treated as conditionally exogenous.

[28] Brückner et al (2012) use oil price shocks as the IV for GDP in testing Wagner's law using Engel curves for public spending. However, in the present context, oil price shocks could induce higher SP spending by altering the distribution of income (invalidating the IV). Instead, we consider ways in which income distribution can be a potential confounding variable in its own right.

We then test an encompassing model, recognizing that these variables are correlated with each other.

4.1 Regression specifications

The following linear regression model of the SPEC is used to test each hypothesis separately:

$$\frac{SP_{it}}{GDP_{it}} = \alpha + \beta lnGDP_{it} + \gamma X_{it} + \varepsilon_{it} \ \ (i = 1, \dots, N;\ t = 1, \dots, T) \qquad (1)$$

Here SP_{it} is social protection spending in country *i* at date *t*, GDP_{it} is GDP per capita, and X_{it} is the control variable implied by each hypothesis. The income elasticity of demand for SP is then given by $1 + \beta/(\frac{SP_{it}}{GDP_{it}})$; an upward slopping SPEC ($\beta > 0$) implies that SP is a luxury good, i.e., with an income elasticity exceeding unity. We consider two assumptions about the error term, ε_{it}, in equation (1). For the first, it is taken to be a standard innovation error term, orthogonal to the regressors, and to selection into the panel data set, while in the second, we include country fixed effects, potentially correlated with the regressors and with selection.

We also estimate a more general (encompassing) specification that generalizes (1) in two respects. First, given that the controls are also correlated with each other, the encompassing model includes all the controls so they can fight it out as to which is more important. (When there is more than one for a given hypothesis, we select what appears to be the best representative.) Second, we allow for any (continuous) nonlinearity in the SPEC. Such nonlinearity in the national income effect on SP spending shares (as evident in Figure 1) may well be confounding in this context. (For example, if the true relationship is quadratic, but we exclude the squared value, then a control variable may simply pick up this omitted variable.) Combining these assumptions, our encompassing regression has the following form:

$$\frac{SP_{it}}{GDP_{it}} = f(lnGDP_{it}) + \boldsymbol{X}_{it}\boldsymbol{\gamma} + \eta_i + \nu_{it} \ \ (i = 1, \dots, N;\ t = 1, \dots, T) \qquad (2)$$

Here $f(.)$ is some (data-determined) smooth nonparametric function and $\boldsymbol{X}_{it}$ is now a vector of control variables. (The income elasticity of demand for SP is then $1 + \frac{\partial lnf(.)}{\partial lnGDP_{it}}$.) We estimate (2) as a partial linear regression (Yatchew 1998) using Stata's PLREG routine (Lokshin 2006).

In correcting for selective reporting, the propensity scores (p-scores hereafter) are estimated using a probit regression for whether SP spending is data for a given country and year. In addition to (log) GDP per capita, the regressors are the log of the country population in that year and a variable measuring the availability of SP spending data for that country in prior years.[29] By including population size our reasoning is that public spending data have properties of a public good, such that lower average costs in more populous countries entail that the data are more likely to be available. The variable for past SP data is motivated by the assumption that past experience in providing SP spending data makes it less costly to currently provide such data. We then include a cubic function of the p-scores in the regressions for the SPEC as control functions for testing the selective reporting hypothesis.[30]

4.2 Results

As can be seen from the "baseline" panel in Table 3, the slope coefficient in the simplest linear regression specification for the SPEC gives a semi-elasticity of 3.78 (implying a full elasticity of SP spending to GDP of 1.42, when evaluated at the mean SP share for 2019). This falls to a semi-elasticity of 1.37 (full elasticity in 2019 of 1.15) when the country effects are included. Qualitatively, this is what one would expect if the omitted country characteristics in the SPEC based on cross-country comparisons (such as Figure 1) tend to be jointly positively correlated with the SP share and GDP. Nonetheless, a positive and significant income effect on the SP share of national income remains when we allow for the latent country effects.

We still find a statistically significant non-linear income effect on the SP shares when we allow for country fixed effects in the PLREG estimator (the "baseline" panel in Table 3). Figure 4 compares the estimated nonparametric sub-functions, $\hat{f}(lnGDP_{it})$, with and without the country effects, where the $\hat{f}(lnGDP_{it})$ values controlling for the country effects are centered on the overall mean of the country-specific means of $\frac{SP_{it}}{GDP_{it}}$. The positive slope is evident either way, but is attenuated considerably by introducing the country effects. Comparing the top quintile of

[29] More precisely, this variable measures the share of previous country-year observations (from 1995 and up to the year previous to that of each observation) where SP spending data is not missing. When data are not missing in every year since 1995 the variable takes a value of one, and when data is always missing the variable takes a value of zero. The variable takes intermediate values whenever information in some years is missing and non-missing in others.
[30] For a more general treatment of these methods of correcting for sample selection bias, see Das et al. (2003).

countries ranked by GDP per capita with the bottom quintile, the mean SP shares of national income are 9% and 5% with country fixed effects, as compared to 13% and 2% without them.

Next, we test the effects of adding time-varying country characteristics, corresponding to the various hypotheses laid out in Sections 2 and 3. The results in testing the hypotheses one at a time (equation 1) are also found in Table 3. This uses total SP spending. The Addendum provides corresponding results obtained by dropping public spending on pensions (Table A4). The main results are robust to this change, though with some (un-surprising) differences in the coefficients, notably that the coefficient on the population share over 65 is much reduced when one excludes spending on pensions. In addition to standard tests, we provide the test described in Lokshin (2006) for whether the nonparametric sub-function estimated using PLREG is significantly different from a constant (i.e., no income effect).

Excluding the country fixed effects, one could claim empirical support in Table 3 for each of the hypotheses outlined in Section 2. Adding the country fixed effects (thus identifying the effects solely from the inter-temporal variances) still leaves support for the Prices, Accountability and Technology Hypotheses. Support for the Governance and Aging Hypotheses is diminished with fixed effects.

In most cases, adding any one of the postulated covariates reduces the GDP slope of the OLS SPEC, though the effect remains statistically significant. Significant income effects in the PLREG estimates are also indicated, in general.

Recall that in testing the selective reporting hypothesis we use three regressors in the probit for the availability of SP spending data, namely log GDP per capita, log population and the dummy variable for whether SP spending data had been already available in the data set. All three had the expected signs and were significant at the 5% level or better; the pseudo R^2 was 0.423. (Details can be found in the Addendum Table A3.) The income effect remains strong with the correction for bias due to selective reporting (Table 3), so this is not why we see a rising SPEC.[31]

[31] Note that only the last of the regressions in Table 3 corrects for sample selection bias, since we treat this as a distinct hypothesis rather than in combination with other hypotheses. The encompassing regression in Table 4 corrects for sample selection as a member of the set of joint hypotheses.

We turn now to the encompassing test (equation 2) in Table 4. Since the sample is very different (including smaller) when we include relative prices, we leave these out for Table 4, but the Addendum includes the regressions with relative prices (Table A5). The results are similar.

In both the linear specifications estimated by OLS (with and without fixed effects), the multiple covariates flatten the SPEC, i.e., the coefficient $\hat{\beta}$ on log GDP in the linear regression is much reduced and no longer significantly different from zero implying an income elasticity of unity. (Table 4 is for total SP spending; the Addendum Table A6 provides the corresponding regressions when one drops pensions; the results are similar.) However, when we allow for nonlinearity using PLREG, we still find a significant income effect. There is support for the Accountability Hypothesis in the specification without country fixed effects, but this switches sign when we introduce those effects; time periods in which accountability worsened saw rising SP spending, possibly to compensate. There is also support for the Domestic Resource Mobilization and the Aging Hypotheses without country fixed effects but this is not robust to allowing for the country effects. The only covariate that remains strong statistically across different specifications in Table 4 is internet usage. However, we warn against too much emphasis on any one factor given the obvious sensitivity of coefficients and their standard errors to the data and econometric specification.

Figure 5 gives the SPEC when we control for all the covariates in Table 4 set at the overall means, with and without country fixed effects. Without country effects, we still see a rising SPEC, though with a greatly attenuated slope relative to Figure 1. Now we see a mean SP share of around 9% of GDP among low-income countries. Strikingly, when we add country effects, the income gradient reverses its sign over most of the range. In marked contrast to the "unconditional" SPEC in Figure 1, we now find that the conditional mean share of GDP going to SP falls from 16% in the poorest countries to around 7% in the richest. While we warn against attributing this dramatic change in the SPEC to any one covariate, we note that redoing Figure 5 only controlling for internet usage at its overall mean (along with country fixed effects) gives a similar picture (Addendum Figure A5). With this one control for time-varying confounders, the SPEC is negatively sloped over most of the income range, falling from a mean share of 12% for the poorest country to 7% for the richest.

5. Testing the hypotheses using data on SP responses to the pandemic

In Section 3.2, we saw that the SPECs have not changed over time in a way that is suggestive of a shift in development priorities in poor and middle-income countries such that more of their national income is devoted to SP. Yes, there has been an increase in the SP share of GDP over time among initially poor countries, but this has been driven by the fact that they are no longer quite so poor. However (as noted above), there may well be a degree of stickiness over time in the SPECs. (Line ministries are known to be resistant to cuts in their budgets.) Possibly public spending decisions simply adjust very slowly to the new priorities.

The COVID shock induced SP responses in varying degrees across the world, as documented by Gentilini et al. (2021). This shock-induced change in policies may better reveal the true change in the underlying policy priorities (even prior to the pandemic), in that the pandemic gave policy makers a new opportunity to implement the policies they truly want—a break from the past.

Studying the pandemic responses might also provide a new perspective on what country-level factors influence the success of pro-SP policy reforms. It may well be that the country characteristics relevant to SP spending in normal times differ from those that matter most in responding to a large shock, or in policy reforms. For example, ICT infrastructure is likely to matter to the scope for scaling up existing programs while general government effectiveness may matter more to the capabilities for introducing new programs, which can be administratively demanding.

We find that poorer countries devoted a smaller share of their GDP to social protection in response to the pandemic. The bold line in Figure 6 plots the relationship.[32] (We explain the dashed line below.) The pattern is similar to Figure 1, although the mean shares of GDP are lower when we focus on the pandemic response alone. Similarly to Figure 1, we still see a marked tendency for the share of GDP devoted to the SP response to rise with GDP per capita. The factor of three or more in the ratio of mean SP share for high-income countries over that for low-income countries is also evident in Figure 6.[33]

---

[32] This is an updated version of a similar graph found in Demirgüç-Kunt et al. (2022) though the latter paper used a linear functional form in representing the line of best fit.
[33] Granted, there are some notable differences; for example, the U.S. is well below the regression line in Figure 1, but well above it in Figure 6. However, taken overall, we do not find that the pandemic-induced policy shock generated a cross-country SPEC that looks very different to what we found prior to the pandemic.

The data on SP spending in response to the pandemic form a single cross-sectional data set, as distinct from the longitudinal (panel) data we assembled for looking at SP spending prior to the pandemic. So, we cannot include country fixed effects. However, as a check for bias associated with latent country characteristics, we include the estimated country effect from the pre-pandemic panel data regressions as an extra control in the encompassing model.[34] Notice that, on *a priori* grounds, the sign on this latent effect could go either way. On the one hand, countries with greater prior experience with SP interventions may have seen lower costs of implementing an extra SP effort during the pandemic. But, on the other hand, they would presumably have less need for extra SP effort during the pandemic.

Similarly to our analysis of the pre-pandemic data, we build up our tests using regression controls to augment Figure 6. Table 5 provides the results in testing the hypotheses one-at-a-time. Relative prices for human development services do not shift the SPEC much. We find mixed support for the Accountability Hypothesis. In the linear regression, the VA and Polyarchy indices have positive coefficients, significant at the 5% level, but these effects are not robust to allowing for nonlinearity using the PLREG estimator. The GDP effect remains strong, however, so we cannot conclude that the Accountability Hypothesis explains that effect.

There is stronger support for the Governance 2 Hypothesis in the SP response to the pandemic. Strikingly, we find that GE knocks out GDP per capita, leaving it with a small effect not significantly different from zero. The GE indicator has a positive and significant effect in both the linear and non-linear SPECs.

We also find support for the Technology Hypothesis, with mobile phone usage having a significant effect at the 5% level; the GDP effect is attenuated somewhat when we control for mobile phones but remains strong. In contrast to the pre-pandemic SP spending data, we find no significant effect of internet usage in the pandemic responses.

There is support for the Ageing Hypothesis in the linear specification, with the population share over 65 having a significant effect at the 5% level, but this is not robust to allowing for nonlinearity in the SPEC.

---

[34] Since this is a generated regressor, we bootstrap the standard errors. We use the country effects estimated in the OLS specification of pre-pandemic data (third column in Table 4).

Table 5 provides the test of whether it is the COVID impact that explains the income effect. We see that COVID mortality is not the reason for the income effect in Figure 6; we do not see higher SP spending by richer countries because they were hit harder by the pandemic.

So the only covariate that, on its own, can robustly account for the income effect of the SP share in pandemic responses is the government effectiveness index. And we find that the GE effect dominates the income effect. The dashed line in Figure 6 plots the relationship with GDP but now with the control for GE (set at the overall mean) using the PLREG estimator to assure flexibility in representing the income effect on the SPEC. Once we control for GE, the share of GDP devoted to SP during the pandemic declines as GDP per capita rises, although statistically one cannot reject the null hypothesis of zero slope (Table 5). The quantitative effect of GE differences is sizeable. Consider the 20 countries with the lowest GDP per capita in our data set. If their GE index was the same as the overall mean index then their share of GDP devoted to SP would have risen from 0.86 to 2.05 percent of GDP.

To this point, we have tested each hypothesis one at a time. Table 6 combines the controls, and also provides the PLREG estimates of equation (2), allowing a flexible representation of the GDP effect. The government effectiveness index not only knocks out the GDP effect, it is also the (statistically) strongest control variable in explaining the SP shares during the pandemic. There is also support for the Ageing Hypothesis in the encompassing model once we allow for latent country effects on the SPEC (based on the retained country effects from the corresponding regressions in Section 4), which have a negative sign.

We have seen that SP spending in response to the pandemic showed a very similar relationship with national income to pre-pandemic SP spending, though the covariates are somewhat different, with a more important role played by government effectiveness, and (much) less by ICT. However, SP was only one component of the policy response to the pandemic; extra spending was also done in health and infrastructure, for example. If the constraints on assuring effective SP in poor countries are not as severe for other (non-SP) pandemic policies then the governments of poorer countries may have been drawn to substitute toward other types of spending in their response to the pandemic.

That conjecture is not consistent with what we see in the data. Figure 7(a) provides the Engel curve for total stimulus spending while 7(b) provides it for non-SP stimulus spending. We see that the pattern is similar during the pandemic (comparing to Figure 1) although the (unconditional) pandemic SPEC is flatter until upper-middle incomes area reached. When we control for government effectiveness, the share of GDP devoted to both total stimulus spending and its non-SP components shows a U-shaped relationship, with relatively high shares implied for poor countries. We infer that the constraints stemming from weak governance applied not only to SP spending, but also impacted non-SP spending responses to the pandemic.

6. Conclusions

Poor countries devote a much lower share of their national income to social protection than rich countries; in other words, we see a rising Engel curve for social protection across countries. This has led many observers to argue that social protection spending is a luxury good. However, declaring something to be a luxury good does not get us very far as an explanation—it is easy enough to say, but if one stops there, one is left in the dark about what made it so. We have assessed whether the popular luxury good interpretation is robust to the existence of confounding variables (correlated with both the share of GDP going to SP and GDP). The main confounders can be interpreted as differences in the costs incurred by governments in supplying social protection.

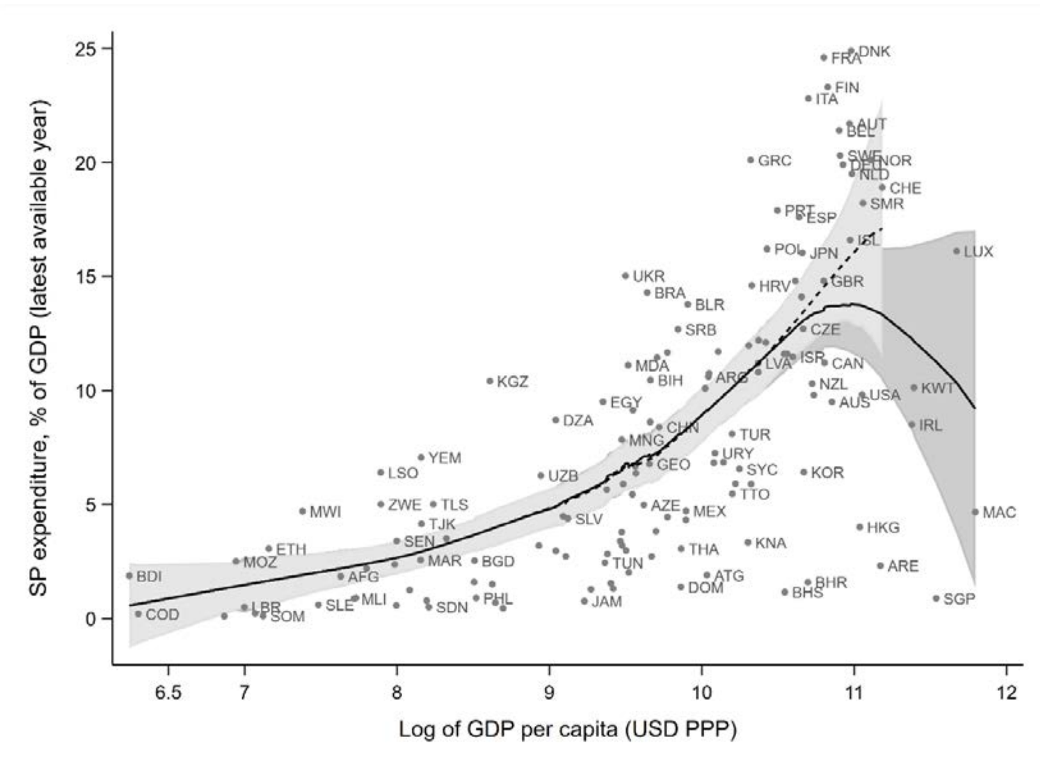
The paper has pointed to a number of alternative hypotheses that are suggestive of potentially confounding time-varying variables. There are undoubtedly other covariates that we have not adequately accounted for. However, we do not need to account for all possible confounders; it is plain from the set of controls we have considered here that the luxury good hypothesis is not robust to allowing for plausible confounders. Indeed, along with country fixed effects, our identified confounders not only reduce the slope of the SP Engel curve, they also change its sign, suggesting that the conditional mean share falls over a wide range of incomes—that SP is more like a necessity than a luxury.

Nor does the fact that richer countries have been observed to devote a higher share of their GDP to social protection during the pandemic imply that social protection is a luxury good even when facing such a large economy-wide shock. Here again, we find that omitted variables are confounding the “luxury good” interpretation. When it comes to implementing SP policy responses to the pandemic, or any other big shock, the effectiveness of government in delivering public services more generally can be expected to be a decisive factor. Scaling up existing SP programs will no doubt play a role, but rapid responses to a shock will often require rapid resource mobilization and the ability to design and implement new policies (with new target beneficiaries, such as those whose employment is at risk), all of which will be easier with greater (pre-pandemic) capabilities for effective public service delivery.

Indeed, we have shown that if one controls only for government effectiveness (set at the global mean), then the share of GDP devoted to social protection during the pandemic is essentially no different comparing rich countries with poor ones. In other words, social protection during the pandemic was neither a necessity nor a luxury. Rather than interpreting low SP spending in response to the pandemic (and indeed low total stimulus spending, including non-SP components) as confirmation for the luxury good hypothesis, our results suggest that this reflects weaknesses in governmental effectiveness generally—weaknesses that are highly correlated with average income, but still have a degree of independent variation.

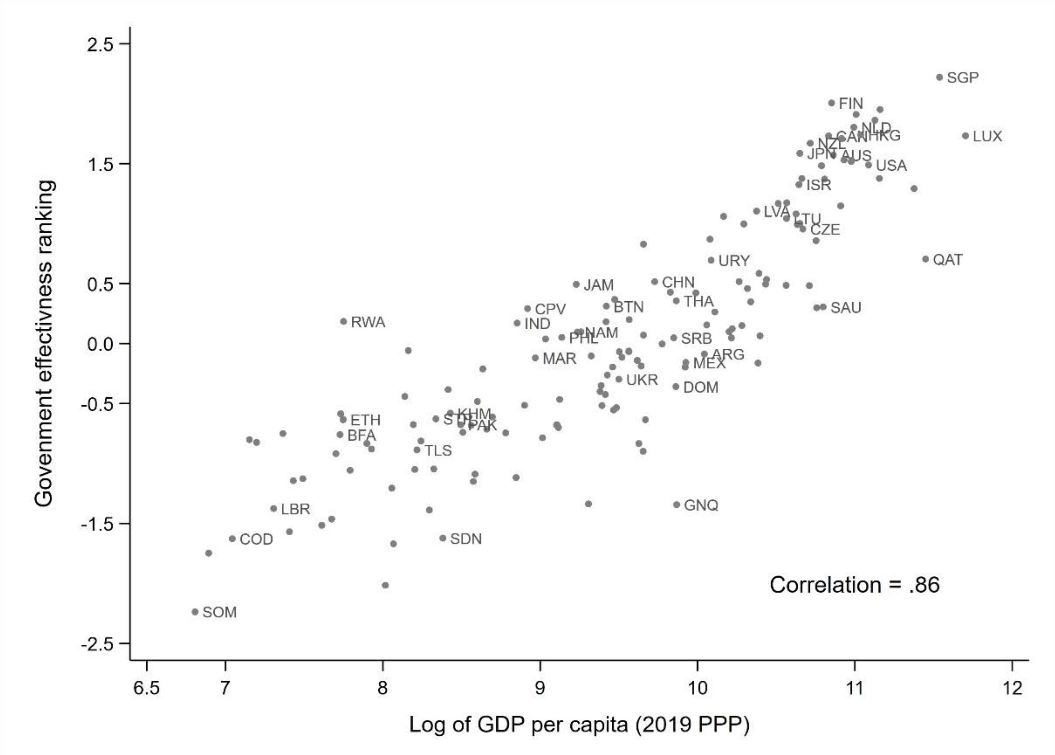
The paper's results point to the importance of exploring further how broader efforts to improve governance and access to technology in developing countries may help attain better social protection. Without success in such efforts, poor countries may be caught in a vicious cycle whereby weakness in these correlates of low income inhibits effective social protection, which helps maintain poverty.

Figure 1: Scatter plot of social protection spending as a share of GDP against log GDP per capita for latest available pre-pandemic year and the nonparametric SPEC



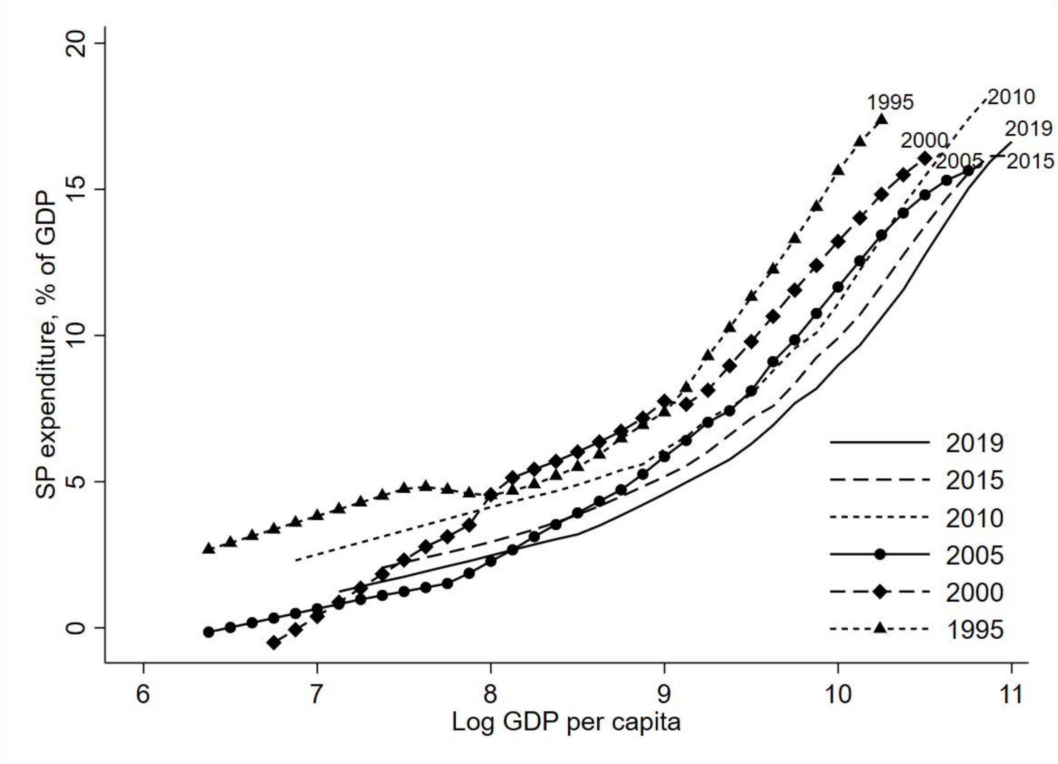
Note: The nonparametric regression line is a smoothed scatter plot (using the lowess command in Stata). The shaded area is the 95% confidence interval. The solid black line includes all the countries and the dashed one excludes the five richest countries as measured by their log GDP per capita (Macao, Luxembourg, Singapore, Kuwait, and Ireland). The Addendum provides the corresponding graph deleting pension spending. Country codes are the U.N.'s Alpha 3 list.

Figure 2: Government effectiveness and GDP per capita



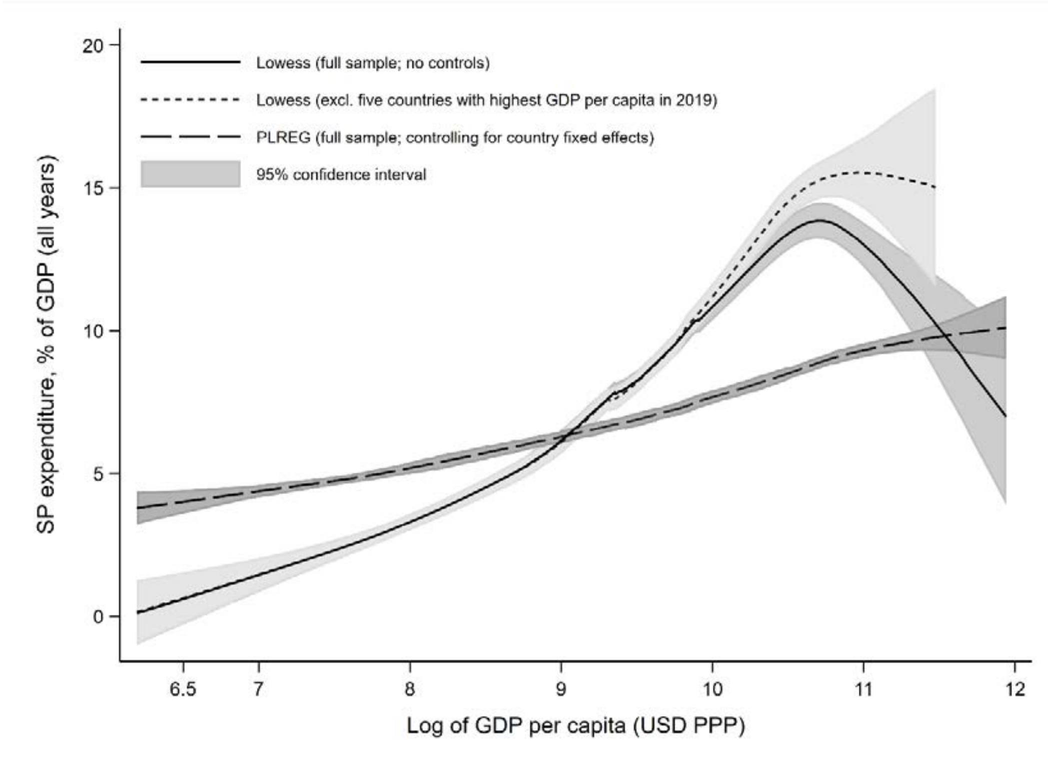
Note: The graph plots, for the 154 countries with the required data, the 2019 WGI government effectiveness index (vertical axis) against the log 2019 GDP per capita, in USD PPP prices (horizontal axis). Country codes are the U.N.'s Alpha 3 list.
Source: Government effectiveness data from WGI (2022). GDP data from WDI (2022).

Figure 3: Nonparametric SPECs for various years



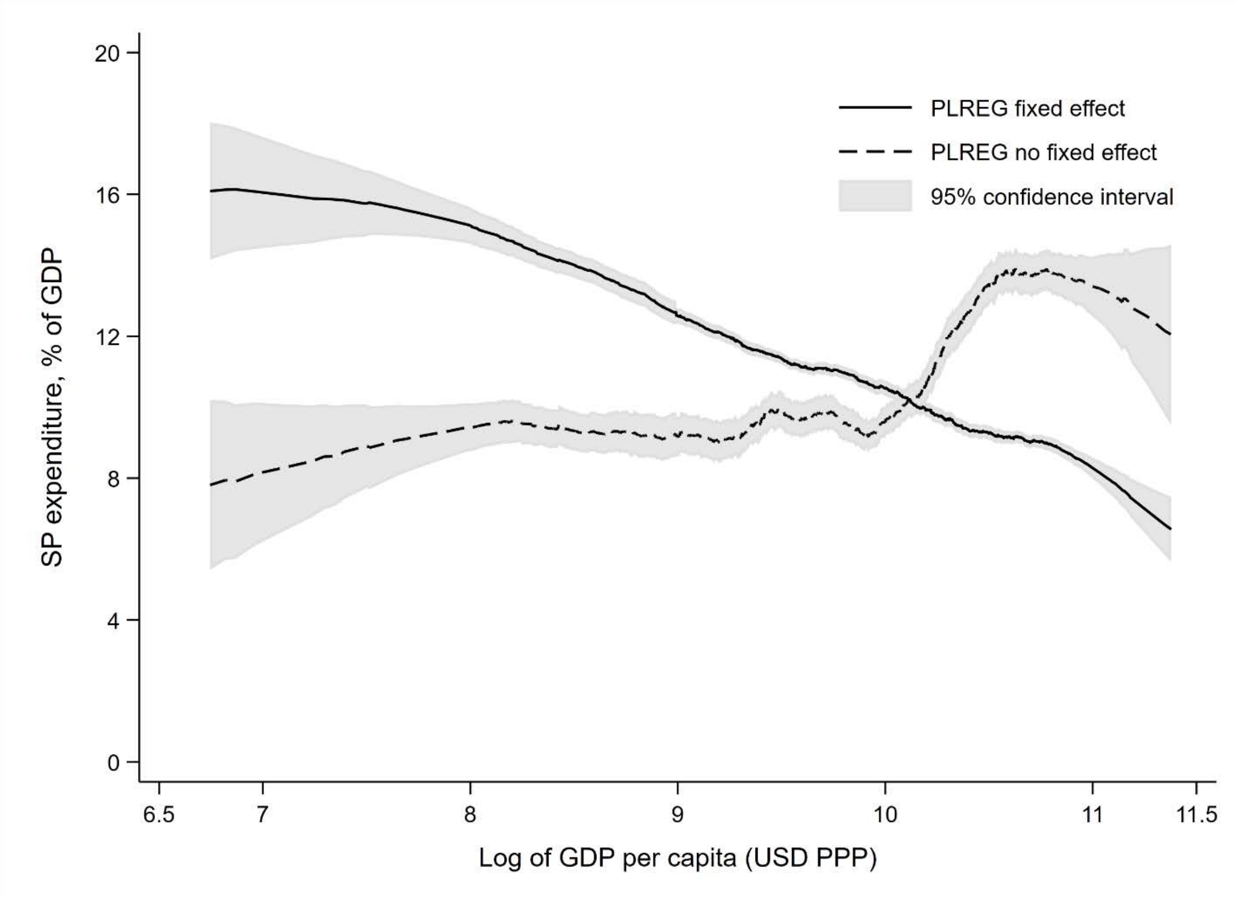
Note: The nonparametric regression lines are smoothed scatter plots (using the lowess command in Stata). Each line represents the data for the specific year indicated. The five richest countries in the income distribution of every year are excluded from each lowess estimation.

Figure 4: Nonparametric SPECs pooling countries and years with and without country fixed effects



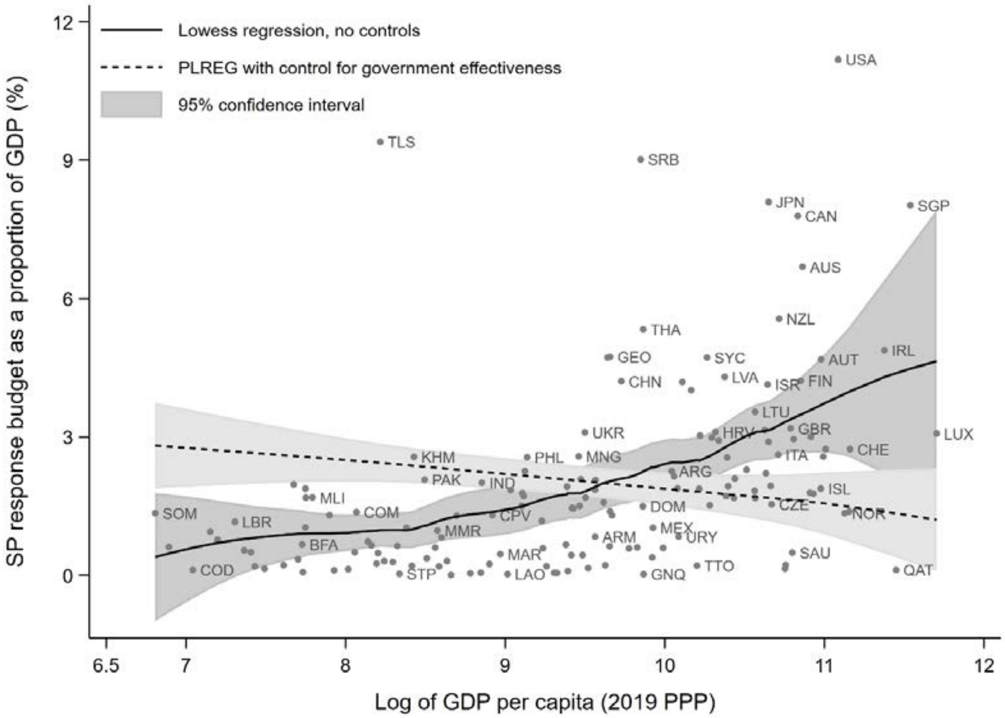
Note: The graph plots social protection spending as a percent of the GDP (vertical axis) against the log GDP per capita, in USD PPP prices (horizontal axis) for the pooled sample of 142 countries over the period from 1995 to 2020. The solid black line is a smoothed scatter plot (using the lowess command in Stata) including all the countries and the short-dash line excludes the five richest countries as measured by their log GDP per capita in 2019 (Macao, Luxembourg, Singapore, Kuwait and Ireland). The (nonparametric) regression line in long dashes includes all countries and controls for country fixed effects. The line is centered on the country mean of the SP share.
Source: Social protection budget data is from the authors' data set. GDP data from WDI (2022).

Figure 5: Nonparametric SPECs pooling countries and years with and without country fixed effects and controlling for all time-varying covariates



Note: The graph plots SP spending as a percent of the GDP (vertical axis) against the log GDP per capita, in USD PPP prices (horizontal axis) for the pooled sample of 142 countries over the period from 1995 to 2020. The (nonparametric) regression line controls for all covariates set their overall mean levels.
Source: Social protection budget data are from the authors' data set (as described in the text and the Addendum). GDP data from WDI (2022).

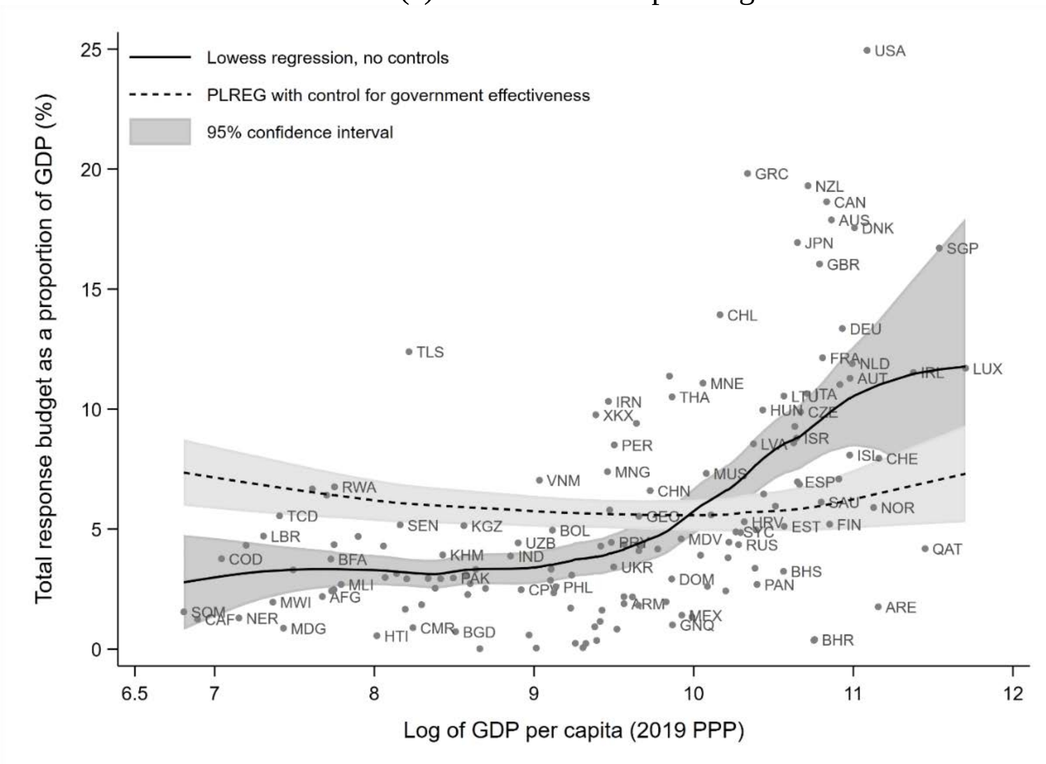
Figure 6: Social Protection Engel Curve during the pandemic



Note: The graph plots, for 154 countries with the required data, the social protection response budget to the COVID-19 pandemic, as a percent of the 2019 GDP (vertical axis) against log GDP per capita for 2019, in USD PPP prices (horizontal axis). The dashed line gives the nonparametric SPEC when one controls for the Government Effectiveness indicator (set at its global mean value).
Source: Social protection budget data is from the authors' data set (described in the main text and the Addendum). GDP data from WDI (2022).

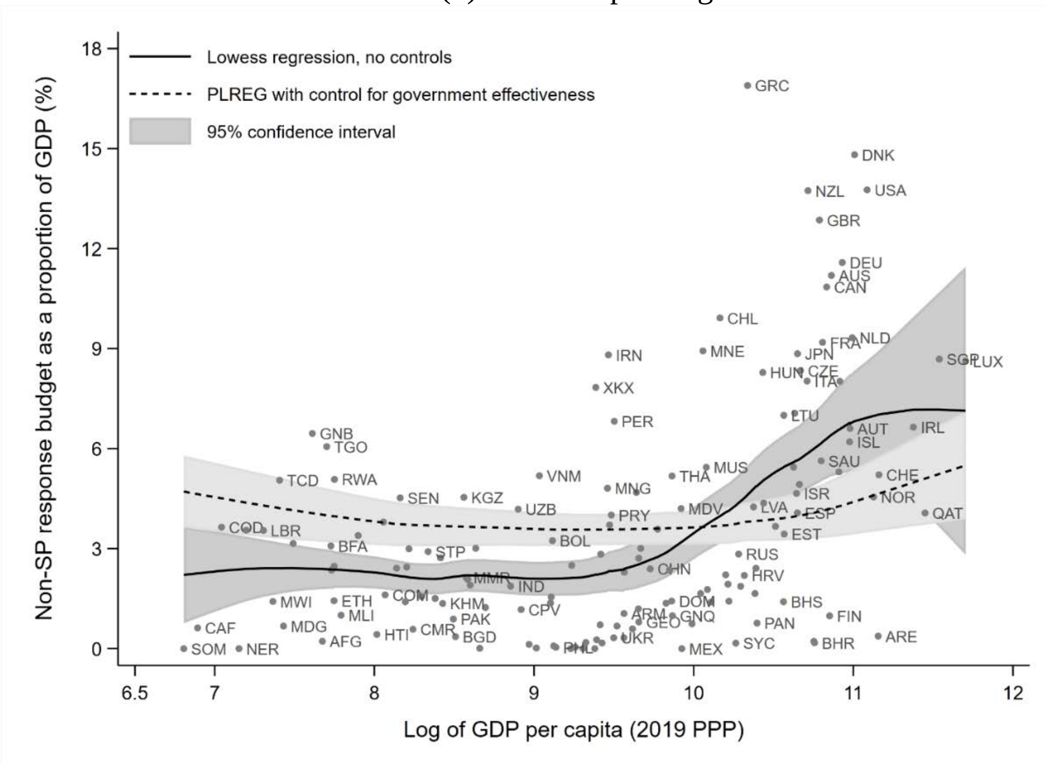
Figure 7: Total stimulus spending and non-SP spending during the pandemic

(a) Total stimulus spending



Note: The graph plots, for 154 countries with the required data, the total response budget to the COVID-19 pandemic, as a percent of the 2019 GDP (vertical axis) against log GDP per capita for 2019, in USD PPP prices (horizontal axis). The dashed line gives the nonparametric SPEC when one controls for the Government Effectiveness indicator (set at its global mean value).
Source: Social protection budget data is from the authors' data set (described in the main text and the Addendum). GDP data from WDI (2022).

(b) Non-SP spending



Note: The graph plots, for 154 countries with the required data, the non-social protection response budget to the COVID-19 pandemic, as a percent of the 2019 GDP (vertical axis) against log GDP per capita for 2019, in USD PPP prices (horizontal axis). The dashed line gives the nonparametric SPEC when one controls for the Government Effectiveness indicator (set at its global mean value).
Source: Social protection budget data is from the authors' data set (described in the main text and the Addendum). GDP data from WDI (2022).

Table 1: Share of SP spending as % of GDP by year stratified by 1995 GDP per capita

GDP per capita by 1995 quintiles	1995	2000	2005	2010	2015	2019
1 (lowest)	1.37	1.68	2.47	3.26	3.52	4.20
2	3.49	6.64	5.19	6.36	5.74	6.01
3	4.57	4.90	4.84	6.75	6.18	6.83
4	9.59	10.00	9.64	10.97	10.24	10.37
5	14.16	12.79	13.17	14.92	15.02	14.72
Total	8.15	8.35	8.07	8.84	8.48	8.91

Source: Authors' calculations from the data set described in Section 3 and the Addendum.

Table 2: Summary of hypotheses for explaining the rising SPEC

		Further testing?	Control variable
Prices		Yes	Relative prices for health and education services
Governance	1: Accountability	Yes	Polyarchy index
	Accountability	Yes	Political competition index
	Accountability	Yes	Voice and accountability index
	2: Government effectiveness	Yes	Government effectiveness index
	3: Resource mobilization	Yes	Tax revenue as a share of GDP
Distribution	1: Median voter	No	
	2: Unequal growth	No	
	3: Relative poverty	No	
Technology		Yes	Mobile phone suscriptions
		Yes	Internet usage
Demographics		Yes	Share of the population age 65+
Selective reporting		Yes	Polynomial in propensity score

Table 3: Separate tests using pre-pandemic pooled SP data

Hypothesis	Estimation method	Log GDP per capita		Test on $f(lnGDP_{it})$	Control variable		Number of observations
		Coefficient	*Std. error*	Value	Coefficient	*Std. error*	
Baseline	OLS	3.778***	*0.499*				2,451
	PLREG			32.360			2,450
	OLS Fixed Effect	1.394***	*0.254*				2,451
	PLREG Fixed effect			5.375			2,450
Prices: Education	OLS	2.465***	*0.571*		4.239***	*1.399*	282
Education/Food price ratio	PLREG			5.997	4.078***	*1.045*	281
	OLS Fixed Effect	1.847***	*0.469*		0.158	*0.593*	282
	PLREG Fixed Effect			9.213	1.576**	*0.740*	281
Prices: Health	OLS	2.358***	*0.591*		7.272***	*2.199*	282
Health/Food price ratio	PLREG			4.651	4.640**	*1.859*	281
	OLS Fixed Effect	1.703***	*0.476*		1.991*	*1.017*	282
	PLREG Fixed Effect			8.182	2.220*	*1.231*	281
Governance: Accountability 1	OLS	2.917***	*0.478*		9.686***	*2.116*	2,342
Polyarchy index	PLREG			17.065	7.948***	*0.616*	2,341
	OLS Fixed Effect	1.402***	*0.259*		-0.845	*1.282*	2,342
	PLREG Fixed Effect			5.451	-1.116	*0.761*	2,341
Governance: Accountability 2	OLS	3.760***	*0.469*		0.712***	*0.208*	2,189
Political competition index	PLREG			29.297	0.505***	*0.055*	2,188
	OLS Fixed Effect	1.454***	*0.267*		0.060	*0.046*	2,189
	PLREG Fixed Effect			6.020	0.055	*0.054*	2,188
Governance: Accountability 3	OLS	2.360***	*0.557*		2.788***	*0.655*	2,091
Voice and Accountability index	PLREG			10.336	1.978***	*0.204*	2,090
	OLS Fixed Effect	1.412***	*0.251*		-0.641*	*0.374*	2,091
	PLREG Fixed Effect			5.185	-0.860***	*0.287*	2,090

Governance: Govt. effectiveness	OLS	1.718***	*0.653*		2.725***	*0.829*	2,082
Government Effectiveness	PLREG			7.041	1.823***	*0.275*	2,081
index	OLS Fixed Effect	1.498***	*0.261*		-0.569	*0.431*	2,082
	PLREG Fixed Effect			5.155	-0.512*	*0.278*	2,081
Governance: Domestic resources	OLS	3.130***	*0.582*		0.288***	*0.069*	2,039
Tax revenues over GDP	PLREG			19.076	0.272***	*0.090*	2,038
	OLS Fixed Effect	1.237***	*0.280*		-0.004	*0.035*	2,039
	PLREG Fixed Effect			3.866	0.018	*0.019*	2,038
Technology 1	OLS	4.269***	*0.564*		-0.014	*0.009*	2,053
Mobile phone suscriptions	PLREG			30.470	-0.006	*0.004*	2,052
per 100 inhabitants	OLS Fixed Effect	0.694	*0.518*		0.010**	*0.004*	2,053
	PLREG Fixed Effect			4.325	0.013**	*0.002*	2,052
Technology 2	OLS	2.375***	*0.718*		0.063***	*0.018*	2,015
Percentage of population	PLREG			11.346	0.035***	*0.008*	2,014
that uses the internet	OLS Fixed Effect	-0.480	*0.653*		0.038***	*0.010*	2,015
	PLREG Fixed Effect			3.451	0.039***	*0.005*	2,014
Demographics	OLS	1.220***	*0.424*		0.790***	*0.069*	2,329
Share of the population	PLREG			4.816	0.721***	*0.024*	2,328
age 65+	OLS Fixed Effect	1.382***	*0.262*		0.027	*0.037*	2,329
	PLREG Fixed Effect			5.174	0.001	*0.042*	2,328
Selective reporting	OLS	3.064***	*0.152*				2,376
	PLREG			18.636			2,375
	OLS Fixed Effect	1.316***	*0.294*				2,376
	PLREG Fixed Effect			5.603			2,375

Note: Standard errors are clustered at the country level. All sub-function tests are significant at $p<0.001$ level. The correction for selective reporting uses a cubic function of the p-score.

Table 4: Encompassing model using the pre-pandemic dataset

	Linear OLS	PLREG	Linear OLS	PLREG
Country fixed effects	No	No	Yes	Yes
Log GDP per capita	0.659		-0.987	
	(0.918)		(0.805)	
Voice and Accountability index	0.694	0.490*	-0.191	-0.553
	(0.807)	(0.250)	(0.404)	(0.354)
Government Effectiveness index	-0.655	-1.330***	-0.574	-0.221
	(1.045)	(0.323)	(0.473)	(0.348)
Tax revenue over GDP	0.165**	0.183***	0.008	0.020
	(0.067)	(0.022)	(0.036)	(0.020)
Percentage of population that uses the internet	0.037**	0.016**	0.045***	0.049***
	(0.017)	(0.008)	(0.012)	(0.006)
Share of the population age 65+	0.689***	0.619***	0.006	-0.006
	(0.084)	(0.033)	(0.033)	(0.044)
Constant	-7.132		21.836***	
	(7.182)		(7.050)	
Test on the non-linear sub-function of log GDP p.c.		3.704		2.226
p-value		0.000		0.013
Observations	1519	1518	1519	1518

Note: Standard errors are shown in parenthesis below the coefficients, clustered at the country level. All specifications include a correction for selective reporting using a cubic function of the propensity score (see Addendum table A3 for the probit regression results used to estimate the propensity scores).

Table 5: Separate tests using SP responses to the pandemic

Hypothesis	Estimation method	Log GDP per capita Coefficient	Log GDP per capita *Std. error*	Test on $f(lnGDP_{it})$	Control variable Coefficient	Control variable *Std. error*	Number of observations
Baseline	OLS	0.840^{***}	*0.148*				154
Relative prices 1							
Health/Food price ratio	OLS	0.820^{***}	*0.182*		0.783	*0.794*	146
	PLREG			2.035††	0.030	*1.233*	146
Relative prices 2							
Education/Food price ratio	OLS	0.798^{***}	*0.183*		0.584	*0.505*	146
	PLREG			1.863††	-0.032	*0.822*	146
Accountability 1							
Polyarchy index	OLS	0.701^{***}	*0.149*		0.061^{**}	*0.029*	141
	PLREG			0.345	0.059	*0.042*	141
Accountability 2							
Political competition index	OLS	0.726^{***}	*0.149*		0.099	*0.062*	141
	PLREG			0.447	0.075	*0.090*	141
Accountability 3							
Voice and Accountability	OLS	0.632^{***}	*0.181*		0.449^{**}	*0.227*	154
Index	PLREG			0.987††	0.359	*0.294*	154
Governance: govt. effectiveness							
Government Effectiveness	OLS	-0.217	*0.278*		1.465^{***}	*0.333*	154
Index	PLREG			0.129	1.617^{***}	*0.454*	154
Governance: domestic resources							
Tax revenue over GDP	OLS	0.713^{***}	*0.175*		0.038	*0.030*	125
	PLREG			0.267	0.026	*0.039*	125
Technology 1							
Mobile phone suscriptions	OLS	0.606^{***}	*0.186*		0.013^{**}	*0.006*	150
per 100 inhabitants	PLREG			1.082†††	0.014^{**}	*0.007*	150
Technology 2							
Percentage of population	OLS	0.934^{***}	*0.395*		-0.004	*0.017*	151
that uses internet	PLREG			0.534	-0.001	*0.020*	151
Aging							
Share of the population	OLS	0.492^{***}	*0.213*		0.105^{**}	*0.045*	147
age 65+	PLREG			0.594†	0.089	*0.060*	147
Covid-19 hypothesis							
Total COVID19 death	OLS	0.931	*0.166*		-0.245	*0.197*	153
per 1,000,000	PLREG			2.643†††	-0.029	*0.265*	153

Note: Standard errors are clustered at the country level. *** indicates that the coefficient is significant at 1% level, ** at 5% level, * at 10% level. ††† indicates the test p-value <0.001, †† p-value <0.01, and † p-value < 0.1.

Table 6: Encompassing regressions for SP responses to the pandemic

	Baseline specification		With imputed fixed effects	
	OLS	*PLREG*	OLS	*PLREG*
Log GDP per capita GDP	-0.024		0.116	
	(0.458)		*(0.598)*	
Voice and accountability	-0.319	-0.370	-0.492	-0.419
	(0.313)	*(0.363)*	*(0.346)*	*(0.385)*
Government effectiveness	1.544***	1.957***	1.591***	1.750***
	(0.480)	*(0.599)*	*(0.532)*	*(0.637)*
Percentage of population that use internet	-0.017	-0.011	-0.020	-0.021
	(0.018)	*(0.021)*	*(0.022)*	*(0.029)*
Share of population over 65 years	0.096*	0.088	0.173***	0.172**
	(0.052)	*(0.064)*	*(0.064)*	*(0.081)*
Covid-19 deaths per million/1000	-0.121	-0.015	-0.108	-0.028
	(0.222)	*(0.292)*	*(0.221)*	*(0.275)*
Imputed country effect			-0.089*	-0.116*
			(0.048)	*(0.060)*
Constant	2.591		0.712	
	(3.624)		*(4.649)*	
Test on log GDP sub-function in PLREG		0.323$^{\alpha}$		0.432$^{\alpha}$
		0.375$^{\beta}$		0.333$^{\beta}$
R^2	0.285		0.329	
Number of countries	147		120	

Note: *** indicates that the coefficient is significant at 1% level, ** - at 5% level, * - at 10% level. Standard errors are shown in parenthesis below the coefficients, clustered at the country level. Bootstrapped standard errors for the specifications using imputed fixed effects.

$^{\alpha}$ Value of the test that Log GDP effect is significant.

$^{\beta}$ Probability that test value is different from zero.

References

Acemoglu, Daron, Simon Johnson and James Robinson, 2005, "Institutions as a Fundamental Cause Of Long-Run Growth," In Philippe Aghion and Steven N. Durlauf (eds) *Handbook of Economic Growth, Volume IA.* Amsterdam: Elsevier.

Alesina, Alberto and Dani Rodrik, 1994, "Distributive Politics and Economic Growth," *Quarterly Journal of Economics* 108: 465-90.

Alesina, Alberto and Francesco Passarelli, 2019, "Loss Aversion in Politics," *American Journal of Political Science*, 63 (4): 936–947

Auteri Monica and Mauro Constantini, 2004, "Is Social Protection a Necessity or a Luxury Good? New Multivariate Cointegration Panel Data Results," *Applied Economics* 36: 1887-1898.

Behrman, Jere R. 2010, "The International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-poverty and Human Resource Investment Conditional Cash," *World Development* 38(10): 1473-1485.

Besley, Timothy, 1995, "Savings, Credit and Insurance," in *Handbook of Development Economics*, Volume III, Edited by Jere Behrman and T.N. Srinivasan. Elsevier Science.

Besley, Timothy, and Torsten Persson, 2011, *Pillars of Prosperity: The Political Economy of Development Clusters*, Princeton: Princeton University Press.

Bird, Richard, 1971, "Wagner's Law of Expanding State Activity," *Public Finance* 26: 1-26.

Brückner, Markus, Alberto Chong, and Mark Gradstein, 2012, "Estimating the Permanent Income Elasticity of Government Expenditures: Evidence on Wagner's Law Based on Oil Price Shocks," *Journal of Public Economics* 96(11-12): 1025-1035.

Center for Systemic Peace, 2020, "Polity5: Political Regime Characteristics and Transitions 1800-2018". Data set available at: https://www.systemicpeace.org/inscrdata.html

Clemente, Jesus, Carmen Marcuello and Antonio Montañes, 2012, "Government Social Spending and GDP: Has there been a Change in Social Policy?" *Applied Economics* 44:22, 2895-2905.

Cornelisse, Peter, and Kees Goudswaard, 2002, "On the Convergence of Social Protection Systems in the European Union," *International Social Security Review* 55(3): 3-17.

Das, Mitali, Whitney Newey, and Francis Vella, 2003, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies* 70(1): 33-58.

Demirgüç-Kunt, Asli, Michael Lokshin and Iván Torre, 2022, "Protect Incomes or Protect Jobs? The Role of Social Policies in Post-pandemic Recovery." Mimeo.

Dincecco, Mark, 2009, "Fiscal Centralization, Limited Government, and Public Revenues in Europe, 1650–1913," *Journal of Economic History* 69(1): 48-103.

Gentilini, Ugo, Almenfi, M., Orton, I., and P. Dale, 2021, "Social Protection and Jobs Responses to Covid-19." Living Paper, World Bank.

Gómez, Rafael, and Pablo Hernández de Cos, 2008, "The Importance of being Mature: the Effect of Demographic Maturation on Global per capita GDP," *Journal of Population Economics* 21: 589–608.

Graham, Carol, 1997, "From Safety Nets to Social Policy: Lessons from the Developing Countries for the Transition Economies," Article, Brookings Institution.

Han, Kangwook, 2021, "Political Budgetary Cycles in Autocratic Redistribution,." *Comparative Political Studies* 55: 727–756.

Hanlon, Joseph, Armando Barrientos and David Hulme, 2010, *Just Give Money to the Poor. The Development Revolution from the Global South.* Sterling, VA: Kumarian Press.

International Monetary Fund (IMF), 2021, "Fiscal Monitor Database of Country Fiscal Measures in Response to the COVID-19 Pandemic," Fiscal Affairs Department, International Monetary Fund, Washington DC.

Jack, William, Adam Ray, and Tavneet Suri, 2013, "Transaction Networks: Evidence from Mobile Money in Kenya," *American Economic Review Papers and Proceedings* 103(3): 356-361.

Jolliffe, Dean, and Espen Prydz, 2016, "Estimating International Poverty Lines from Comparable National Thresholds," *Journal of Economic Inequality* 14(2): 185-198.

Kristov, Lorenzo, Peter Lindert and Robert McClelland, 1992, "Pressure Groups and Redistribution," *Journal of Public Economics* 48(2): 135-63.

Leser, C. E. V., 1963, "Forms of Engel Functions," *Econometrica* 31: 694- 703.

Lindert, Peter, 1994, "The Rise of Social Spending, 1880-1930," *Explorations in Economic History* 31(1): 1-37.

Lipton, Michael, and Martin Ravallion, 1995, "Poverty and Policy," in Jere Behrman and T.N. Srinivasan (eds) *Handbook of Development Economics, Volume 3,* Amsterdam: North-Holland.

Lokshin, Michael, 2006, "Difference–Based Semiparametric Estimation of Partial Linear Regression Models," *Stata Journal* 3: 377-383.

Meltzer, Allan, and Scott Richard, 1981, "A Rational Theory of the Size of Government," *Journal of Political Economy* 89(5): 914-27.

Milanovic, Branko, 2016, *Global Inequality: A New Approach for the Age of Globalization.* Cambridge, Mass: Harvard University Press.

Mulligan, Casey, Ricard Gil and Xavier Sala-i-Martin, 2004, "Do Democracies Have Different Public Policies than Nondemocracies?" *Journal of Economic Perspectives* 18(1): 51-74.

Peracchi, Franco, 2001, "Patterns of Social Spending in Western Europe," Tor Vergata University Rome, Italy.

Preston, Samuel, 1975, "The Changing Relationship Between Mortality and Level of Economic Development," *Population Studies* 29: 231-248.

Ravallion, Martin, 2006, "Transfers and Safety Nets in Poor Countries: Revisiting the Trade-Offs and Policy Options," in Abhijit Banerjee, Roland Benabou and Dilip Mookerjee (eds), *Understanding Poverty*, Oxford: Oxford University Press.

_______________, 2016a, *The Economics of Poverty: History, Measurement and Policy.* New York: Oxford University Press.

_______________, 2016b, "Toward Better Global Poverty Measures," *Journal of Economic Inequality* 14: 227-248.

________________, 2017, *Interventions against Poverty in Poor Places*, 20th Annual WIDER Lecture, , Helsinki, Finland: World Institute of Development Economics Research.

_______________, 2021, "What Might Explain Today's Conflicting Narratives on Global Inequality?" in *Inequality in the Developing World*, edited by C. Gradin, M. Leibbrandt, and F. Tarp. Oxford: Oxford University Press.

Ravallion, Martin, and Shaohua Chen, 2019, "Global Poverty Measurement when Relative Income Matters," *Journal of Public Economics* 177: 1-13.

Ravallion, Martin, Shaohua Chen and Prem Sangraula, 2009, "Dollar a Day Revisited," *World Bank Economic Review* 23(2):163-184.

Ritchie, Hannah, and Max Roser, 2019, "Age Structure," Our World of Data, Oxford University.

Sabates-Wheeler, R., and S. Devereux, 2008, "Transformative Social Protection: The Currency of Social Justice." In: Barrientos, A., Hulme, D. (eds) *Social Protection for the Poor and Poorest.* Palgrave Studies in Development. Palgrave Macmillan, London.

Shelton, Cameron, 2007, "The Size and Composition of Government Expenditure," *Journal of Public Economics* 91: 2230-2260.

Vella, Francis, 1998, "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources* 33(1): 127-169.

Williamson, John, 1990, "What Washington Means by Policy Reform," in Williamson, John (ed.): *Latin American Readjustment: How Much has Happened*, Washington: Institute for International Economics.

Working, H., 1943, "Statistical Laws of Family Expenditures," *Journal of the American Statistical Association* 38: 43-56.

World Bank, 1990, *World Development Report: Poverty*. New York: Oxford University Press.

World Bank, 2001, *Social Protection Sector Strategy. From Safety Net to Springboard.* Washington DC: World Bank.

World Bank, 2006, *World Development Report: Equity and Development.* Washington DC: World Bank.

World Bank, 2022b, *Worldwide Governance Indicators.* Washington, DC: World Bank

World Bank, 2022c, "G2Px: Digitizing Government-to-Person Payments," Washington, DC: World Bank.

Yatchew, A. 1998, "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature* 36: 669–721.

Yu, Xiaoqing, 2016, "Social Protection Tools for the 21st Century," Voices: World Bank Blog.

Addendum

A1: Data sources on SP spending

Our dataset on social protection expenditures covers 142 countries over the period 1995-2020. We define social protection expenditure as the public expenditure covering social assistance, social insurance (including pensions), and active labor market policies. The dataset contains 2,481 country-year observations coming from the sources listed in Table A1.

For countries in the European Union, including the United Kingdom, we use data from Eurostat's SP expenditure database (table code: *spr_exp_gdp*). To be consistent with our definition of SP, we exclude the expenditure under ESSPROS function "sickness and healthcare". These data cover the whole period 1996-2020 for EU-15 countries, and for most of the period for the remaining EU member states. In total, we use Eurostat data for 587 country-year observations.

For countries in Europe and Central Asia not covered by Eurostat data, we use the World Bank's Social Protection Expenditure and Evaluation Database (SPEED), which covers 27 countries in that region (including Turkey) from 2000 to 2020. In total, we use SPEED for 380 country-year observations.

For OECD countries outside Europe, we use data from OECD's Social Expenditure database. This database provides us information for 11 countries in the period 2000 to 2019. In total we use OECD data for 73 country-year observations.

For countries in Latin America and the Caribbean not included in the OECD's Social Expenditure database, we use data from the Economic Development Division of the Economic Commission for Latin America and the Caribbean (ECLAC). These data correspond to government expenditure classified by function (COFOG). When available, we use the values corresponding to the SP expenditure by the general government; otherwise, we use the values corresponding to the SP expenditure by the central government. This source provides us information for 32countries in the period 1995-2020. In total, we use ECLAC data for 443 country-year observations.

For countries in the remaining regions of the world—Africa, Middle East, South Asia and East Asia and the Pacific—or that are not covered by any of the above databases, we rely on two sources: the IMF's Government Finance Statistics and the World Bank's Public Expenditure Reviews.

The IMF's Government Finance Statistics provide data on government expenditure classified by function (COFOG). When available, we use the values corresponding to the SP expenditure by the general government; otherwise, we use the values corresponding to the SP expenditure by the central government including social security funds. This source provides us information for 94 countries in the period 1995-2020. In total, we use GFS data for 782 country-year observations.

The World Bank Public Expenditure Reviews (PERs) are a series of non-periodical reports analyzing the public expenditure of World Bank client countries. These reports are occasional and do not follow a fixed outline (differently to the IMF's Article IV reviews). Most of these reports contain information on SP expenditure. We use information coming from different PERs for 26 countries. In total, we use PER data for 179 country-year observations.

Lastly, for Benin, Mali and Tunisia we use data coming from the World Bank's BOOST Open Budgets Portal. This portal presents detailed budget information for these countries, which allows us to calculate public expenditure on SP. In total, we use BOOST for 37 country-year observations.

Table A1: Sources for SP expenditure

Source	Country-year observations	Number of countries	Period	Regions
Eurostat	587	28	1996-2019	European Union + United Kingdom
SPEED	380	27	2000-2020	Europe and Central Asia + Turkey
OECD	73	11	2000-2019	Non-European OECD countries
ECLAC	443	32	1995-2020	Latin America and the Caribbean
GFS (IMF)	782	94	1995-2020	Africa, Middle East, South Asia, East Asia and the Pacific + countries not included in above sources
World Bank PERs	179	26	1995-2020	Africa, Middle East, South Asia, East Asia and the Pacific
World Bank BOOST	37	3	2004-2018	Benin, Mali, and Tunisia
Total	2,481	142	1995-2020	

Table A2: Share of SP (non-pension) as % of GDP by year stratified by 1996 GDP per capita

GDP per capita 1996 quintiles	2005	2010	2015	2019
1	0.89	1.30	1.27	0.95
2	2.31	2.10	2.33	1.71
3	2.26	3.53	3.00	4.55
4	5.38	6.67	5.01	4.81
5	8.39	8.86	8.45	8.28
Total	4.64	4.98	4.24	4.54

Note: Too few observations of pension spending in 1996 and 2000 (esp., for lowest two quintiles by GDP) to be considered reliable.

Table A3: Probability of having non-missing SP data. Binary probit estimation.

Variable	Coefficient	Standard Error
Log GDP per capita GDP	0.115***	*0.028*
Log population size	0.027**	*0.014*
Available in the previous year	2.572***	*0.089*
Constant	-2.288***	*0.366*
Pseudo R^2	0.423	
Observations	3,451	

Table A4: Separate hypotheses tests on SPECs using pre-pandemic (non-pension) data.

	Dep var: non pension SP spending (in % of GDP)	OLS (1)	PLREG (2)	OLS (3)	PLREG (4)
	Country fixed effects	No	No	Yes	Yes
1	*Baseline*				
	Log GDP per capita	3.157***		0.485	
		(0.313)		(0.335)	
	Test on log GDP sub-function		33.396		4.601
	p-value		0.000		0.000
	Observations	1384	1383	1384	1383
2	*Prices: Education*				
	Log GDP per capita	1.476***		0.396	
		(0.438)		(0.482)	
	Education/food price ratio	3.561***	3.189***	-0.225	0.190
		(0.924)	(0.731)	(0.501)	(0.780)
	Test on log GDP sub-function		1.528		5.863
	p-value		0.063		0.000
	Observations	192	191	192	191
3	*Prices: Health*				
	Log GDP per capita	1.553***		0.399	
		(0.482)		(0.479)	
	Health/food price ratio	5.944***	4.487***	1.512	1.492
		(1.427)	(1.176)	(1.921)	(1.330)
	Test on log GDP sub-function		1.398		5.874
	p-value		0.081		0.000
	Observations	192	191	192	191
4	*Accountability hypothesis 1*				
	Log GDP per capita	2.219***		0.460	
		(0.278)		(0.313)	
	Polyarchy index	5.967***	5.749***	-1.632	-0.653
		(1.023)	(0.541)	(2.422)	(1.087)
	Test on log GDP sub-function		15.386		4.132
	p-value		0.000		0.000
	Observations	1328	1327	1328	1327
5	*Accountability hypothesis 2*				
	Log GDP per capita	2.828***		0.516	
		(0.299)		(0.342)	
	Political competition index	0.368***	0.316***	0.040	0.087
		(0.083)	(0.047)	(0.044)	(0.059)
	Test on log GDP sub-function		26.149		4.912
	p-value		0.000		0.000
	Observations	1264	1263	1264	1263
6	*Accountability hypothesis 3*				
	Log GDP per capita	1.706***		0.531	
		(0.320)		(0.338)	
	Voice and Accountability index	1.981***	1.951***	-0.383	-0.620
		(0.344)	(0.165)	(0.334)	(0.392)

	Test on log GDP sub-function		7.968		3.843
	p-value		0.000		0.000
	Observations	1227	1226	1227	1226
7	*Governance: govt. effectiveness*				
	Log GDP per capita	0.569		0.602*	
		(0.399)		(0.359)	
	Government Effectiveness index	2.755***	2.575***	-0.631	-1.176***
		(0.419)	(0.177)	(0.501)	(0.284)
	Test on log GDP sub-function		2.154		4.502
	p-value		0.016		0.000
	Observations	1223	1222	1223	1222
8	*Governance: domestic resources*				
	Log GDP per capita	2.935***		0.157	
		(0.362)		(0.285)	
	Tax revenue over GDP	0.139**	0.133***	0.014	0.014
		(0.063)	(0.016)	(0.021)	(0.014)
	Test on log GDP sub-function		19.527		2.502
	p-value		0.000		0.006
	Observations	1157	1156	1157	1156
9	*Technology hypothesis 1*				
	Log GDP per capita	3.596***		-0.893	
		(0.449)		(0.568)	
	Mobile phone subscriptions x 100 inh.	-0.014**	-0.009***	0.016**	0.015***
		(0.006)	(0.003)	(0.007)	(0.003)
	Test on log GDP sub-function		34.985		3.190
	p-value		0.000		0.001
	Observations	1222	1221	1222	1221
10	*Technology hypothesis 2*				
	Log GDP per capita	3.260***		-0.831	
		(0.518)		(0.741)	
	Percentage of population that uses internet	-0.000	-0.011*	0.022*	0.031***
		(0.012)	(0.006)	(0.013)	(0.005)
	Test on log GDP sub-function		19.438		4.339
	p-value		0.000		0.000
	Observations	1202	1201	1202	1201
11	*Ageing hypothesis*				
	Log GDP per capita	2.355***		0.484	
		(0.323)		(0.337)	
	Share of the population age 65+	0.226***	0.200***	0.020	0.050
		(0.061)	(0.021)	(0.037)	(0.037)
	Test on log GDP sub-function		16.615		4.358
	p-value		0.000		0.000
	Observations	1318	1317	1318	1317
12	*Selection hypothesis*				
	Log GDP per capita	2.674***		0.475	
		(0.313)		(0.423)	
	Test on log GDP sub-function		18.550		5.028
	p-value		0.000		0.000
	Observations	1370	1369	1370	1369

Note: Standard errors are clustered at the country level in columns 1 and 3.
*** indicates that the coefficient is significant at 1% level, ** at 5% level, * at 10% level

Table A5: Encompassing model using the pre-pandemic dataset, including the relative price indicators.

Dep var: SP spending (in % of GDP)	Linear OLS	PLREG	Linear OLS	PLREG
Country fixed effects	No	No	Yes	Yes
Log GDP per capita	-0.388		-2.270	
	(1.023)		(1.395)	
Education/Food price ratio	2.731*	3.906***	0.502	1.728*
	(1.593)	(1.212)	(0.776)	(0.904)
Health/Food price ratio	2.953	0.458	1.759	1.893
	(2.136)	(2.243)	(1.119)	(1.378)
Voice and Accountability index	0.447	0.211	-1.288	-1.739**
	(0.787)	(0.646)	(1.030)	(0.849)
Government Effectiveness index	-1.511	-2.013**	-0.088	-1.058
	(1.147)	(0.818)	(0.711)	(0.710)
Tax revenue over GDP	0.209***	0.163***	-0.006	-0.021
	(0.074)	(0.051)	(0.043)	(0.036)
Percentage of population that uses the internet	0.042**	0.021	0.056***	0.044***
	(0.021)	(0.023)	(0.021)	(0.015)
Share of the population age 65+	0.674***	0.669***	0.087	0.174
	(0.084)	(0.077)	(0.101)	(0.113)
				1.728*
Constant	2.634		28.613**	
	(8.277)		(12.444)	
Test on the non-linear sub-function of log GDP p.c.		1.581		4.918
p-value		0.057		0.000
Observations	220	219	220	219

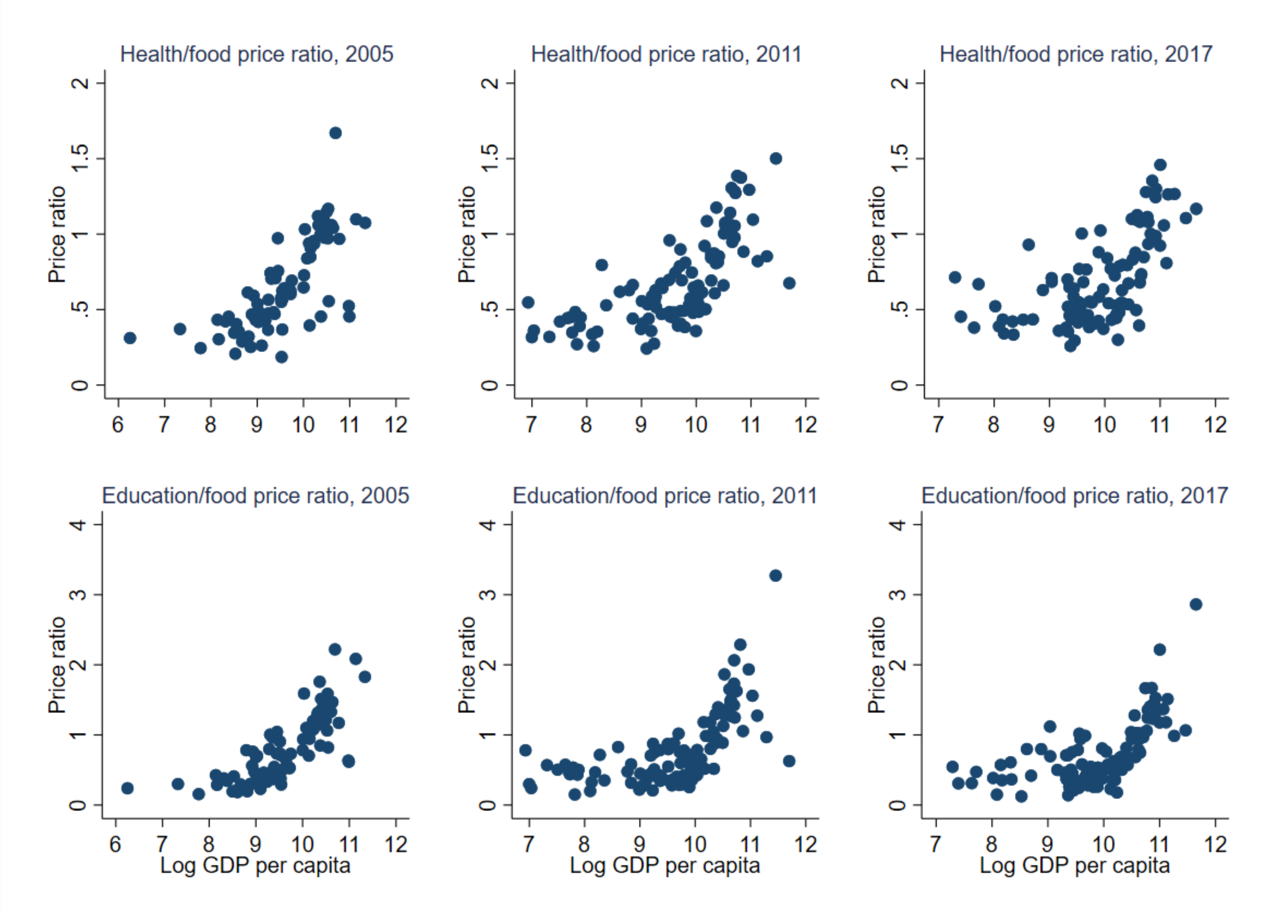
Note: Standard errors are shown in parenthesis below the coefficients, clustered at the country level. All specifications include a correction for selective reporting using a cubic function of the propensity score (see Addendum table A3 for the probit regression results used to estimate the propensity scores).

Table A6: Encompassing model using the pre-pandemic dataset, excluding pensions

Dep var: non pension SP spending (in % of GDP)	Linear OLS	PLREG	Linear OLS	PLREG
Country fixed effects	No	No	Yes	Yes
Log GDP per capita	0.958 (0.662)		-0.821 (0.848)	
Voice and Accountability index	0.564 (0.672)	0.476* (0.270)	-0.341 (0.417)	-0.438 (0.421)
Government Effectiveness index	1.530* (0.792)	1.294*** (0.300)	-0.262 (0.426)	-0.583* (0.325)
Tax revenue over GDP	0.050 (0.047)	0.058*** (0.017)	0.015 (0.020)	0.012 (0.015)
Percentage of population that uses the internet	0.003 (0.013)	0.001 (0.006)	0.019* (0.011)	0.021*** (0.005)
Share of the population age 65+	0.079 (0.074)	0.067*** (0.026)	0.015 (0.034)	0.061* (0.033)
Constant	-4.991 (7.650)		23.980* (12.512)	
Test on the non-linear sub-function of log GDP p.c.		2.844		3.151
p-value		0.002		0.001
Observations	933	932	933	932

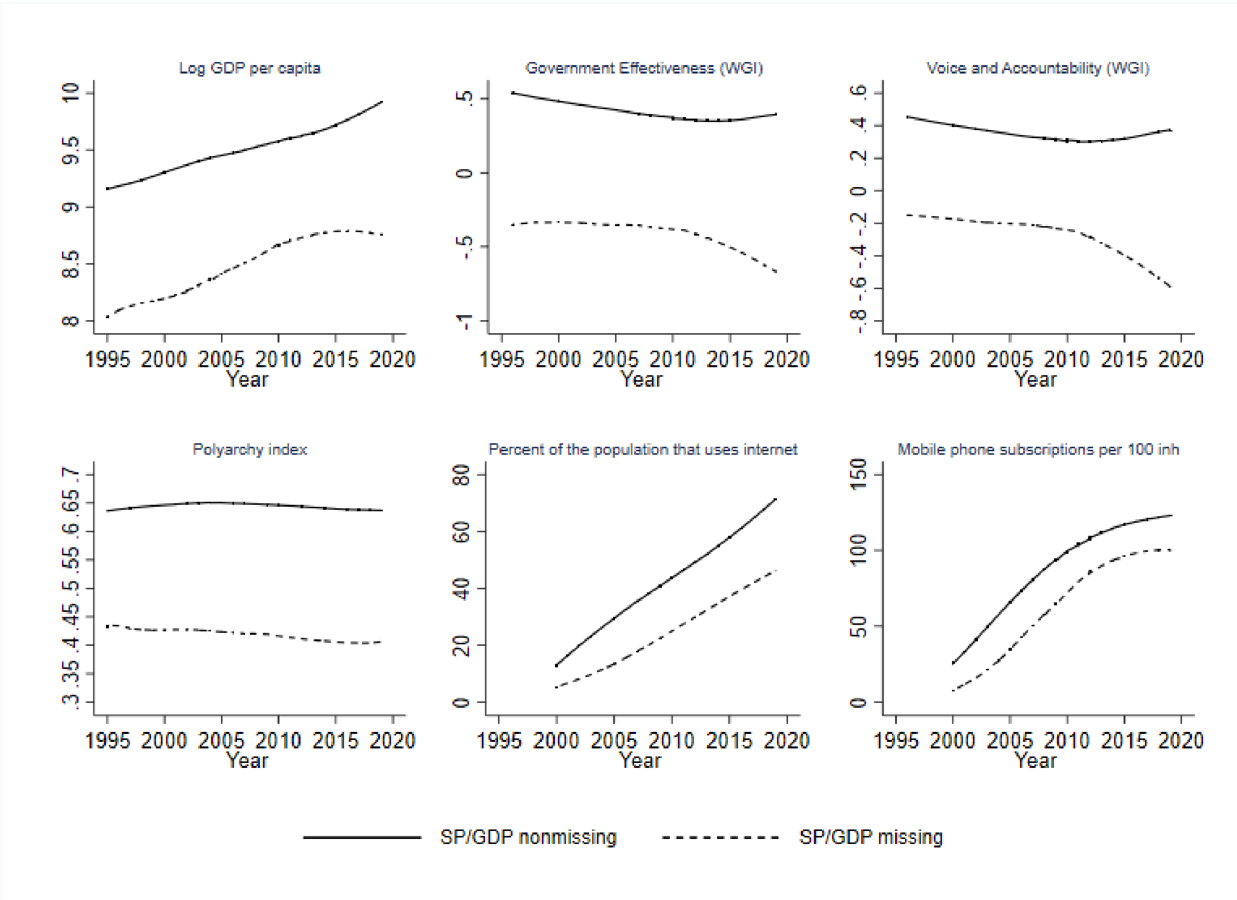
Note: Standard errors are shown in parenthesis below the coefficients, clustered at the country level.

Figure A1: Relative prices of health and education plotted against GDP per capita



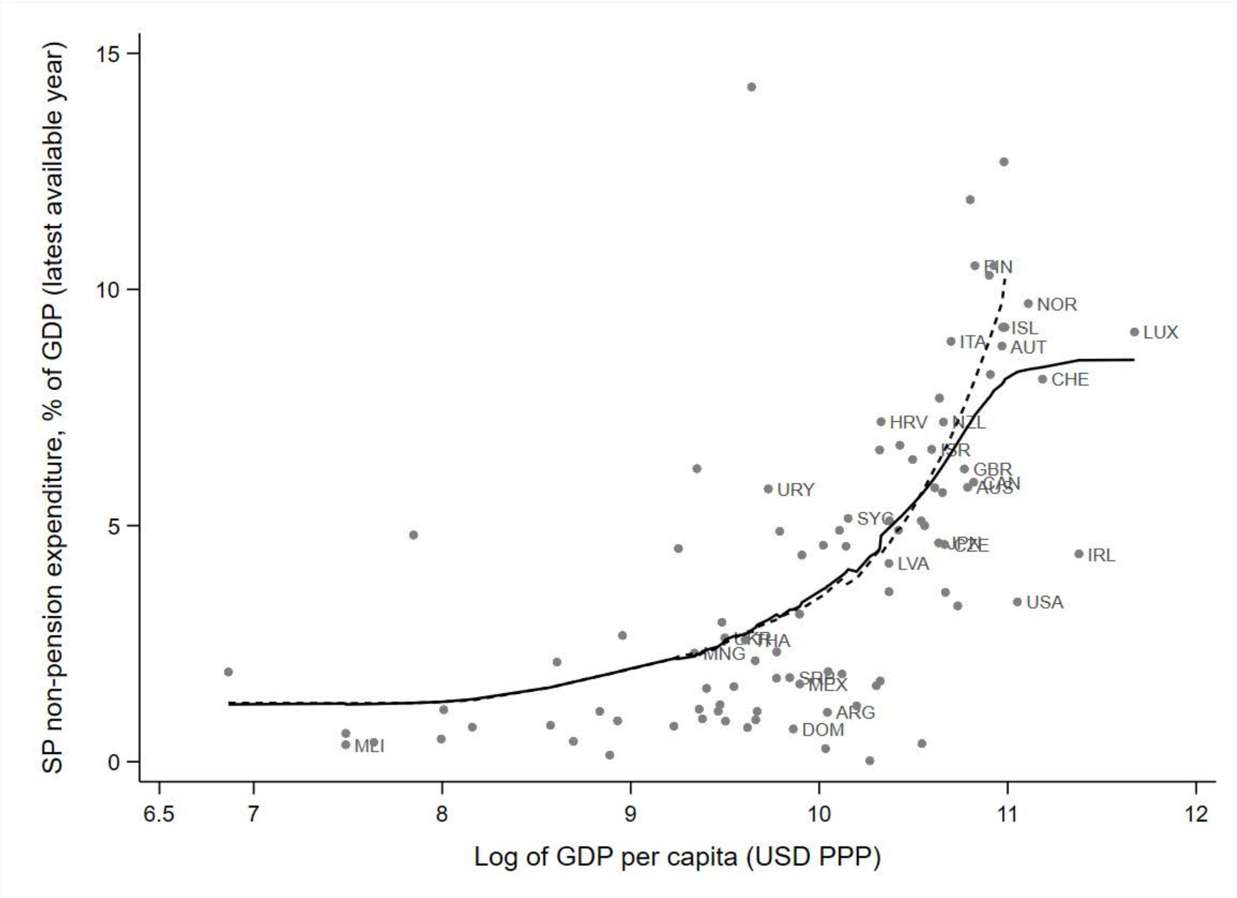
Note: This figure plots the relationship between the log GDP per capita (horizontal axis) and the relative prices of health (vertical axis, top row) and education (vertical axis, bottom row). The relative prices of health and education are expressed with respect to the price of food and non-alcoholic beverages.

Figure A2: Comparisons for the main variables between country-year observations which have SP data available and those that do not



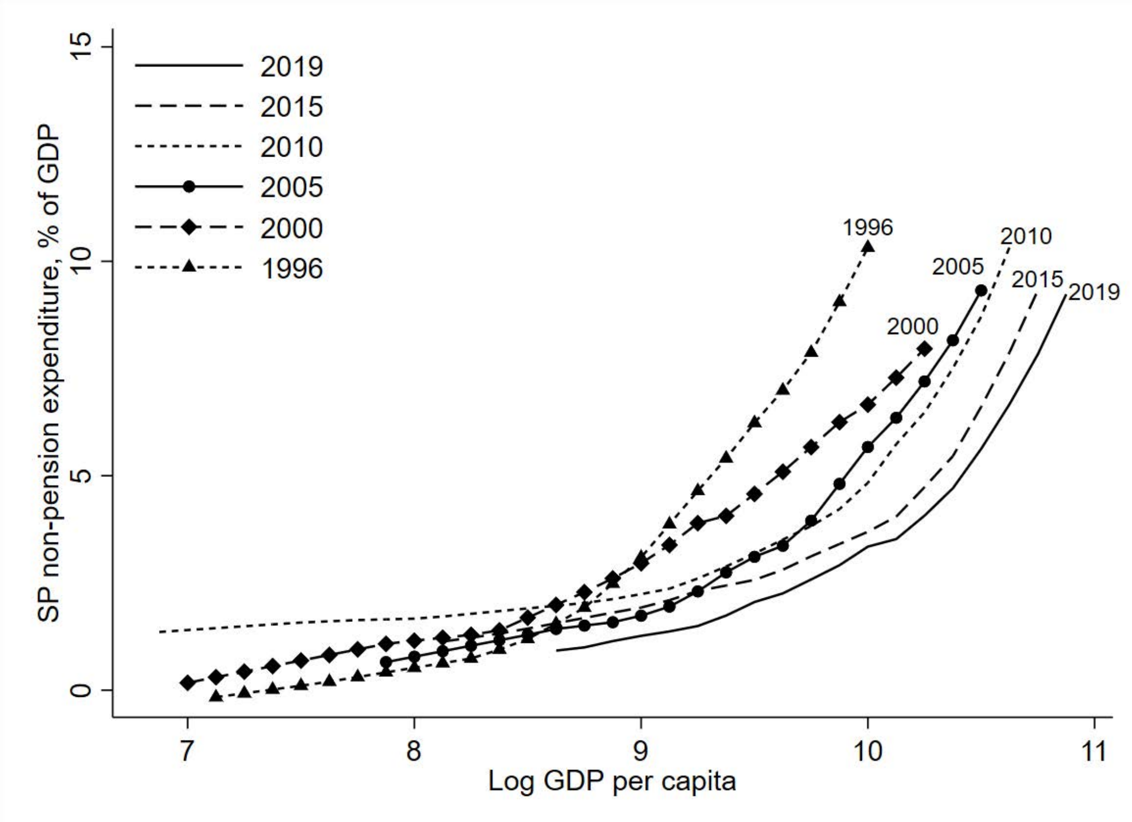
Note: This figure plots the smoothed average over time (using the "lowess" command in Stata) of different covariates (one in each panel) for those country-year observations for which SP/GDP data is available (solid line) and for those country-year observations for which SP/GDP data is missing (dashed line).

Figure A3: Scatter plot of social protection non-pension spending as a share of GDP against log GDP per capita for latest available pre-pandemic year and the non-parametric SPEC



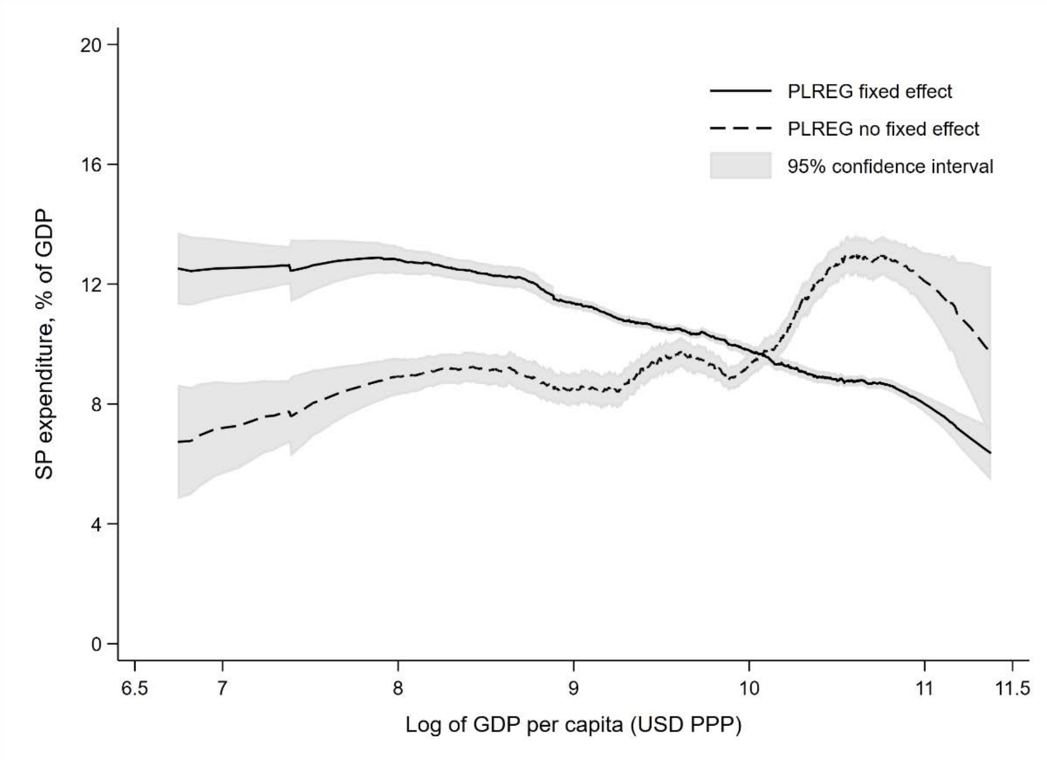
Note: The non-parametric regression line is a smoothed scatter plot (using the "lowess" command in Stata). The solid black line includes all the countries and the dashed one excludes the five richest countries as measured by their log GDP per capita (Macao, Luxembourg, Singapore, Kuwait and Ireland).

Figure A4: Nonparametric SPECs for various years (non-pension spending)



Note: The non-parametric regression lines are a smoothed scatter plot (using the "lowess" command in Stata). Each line represents the data from a specific year. The five richest countries in the income distribution of every year are excluded from each lowess estimation.

Figure A5: Nonparametric SPEC pooling countries and years with and without country fixed effects and controlling for internet usage



Note: The graph plots SP spending as a percent of the GDP (vertical axis) against the log GDP per capita, in USD PPP prices (horizontal axis) for the pooled sample of 154 countries over the period from 1995 to 2019. The (nonparametric) regression line controls for internet usage at the overall mean level and country fixed effects.

Source: Social protection budget data are from the authors' data set (as described in the text and the Addendum). GDP data from WDI (2022).