How Feasible are the Sustainable Development Goals?

Analyzing Government Spending and Policy Priorities Worldwide

Omar A. Guerrero[1,2] and Gonzalo Castañeda[3]

[1]Department of Economics, UCL, United Kingdom

[2]The Alan Turing Institute, United Kingdom

[3]Centro de Investigación y Docencia Económica (CIDE), Mexico

Abstract

We study the feasibility of the Sustainable Development Goals (SDGs) across 140 countries using 79 environmental and socioeconomic indicators provided by the 2020 Sustainable Development Report. We use novel computational methods to estimate the gap that these indicators will have in 2030 in relation to the goals established by the 2030 Agenda. This approach takes into account country-specific networks of interlinkages between the SDGs, heterogeneous levels for public expenditure, and public governance mechanisms that influence how efficient is public spending in improving these indicators. Our results suggest that, even under increased public expenditure, the 2030 Agenda is unfeasible because long-term structural bottlenecks remain in place. We report on the heterogeneity of these structural bottlenecks across countries and SDGs. We argue that these challenges cannot be addressed through sheer spending, but require novel micro-policies intended to affect behaviors, technologies, and organizational practices. One particular set of bottlenecks that stands out relates to the environmental issues contained in SGDs 14 and 15.

The 2030 Agenda for Sustainable Development is a commendable effort by the United Nations to establish a multidimensional and systemic vision on the challenges that humankind must face in the years to come. On the one hand, it recognizes that economic growth should come hand in hand with good working conditions, generalized access to high quality health services, socioeconomic inclusion, lowering gender and income inequalities, a clean environment, the sustainable use of natural resources, physical and legal security for all, the reinforcement of political freedoms, and citizens' participation in public affairs. On the other hand, it argues that all these dimensions of development are intertwined through a complex web of interlinkages and, hence, when designing public policies these issues should not be analyzed in isolation from each other. In spite of this effort, whether governments will be able to achieve these goals by 2030 is a conundrum that has not been properly addressed. Thus, generating quantitative insights on how feasible is the 2030 Agenda and on the role of government expenditure is extremely valuable for the overall planning of development. This is especially important after the economic slowdown produced by the COVID-19 global pandemic of 2020, which will affect the amount and allocation of public spending over the next decade.

In this paper, we study the feasibility of the Sustainable Development Goals (SDGs) across 140 countries. We say that this agenda is entirely feasible for a country when all its development indicators can close the gaps with respect to their goals by 2030. We quantify such goals (specific values for the corresponding indicators) using information presented in the latest edition of the Sustainable Development Report (SDR) [17].[1] Then, we use novel computational methods to estimate the expected gaps that the different SDG indicators would have in 2030, and how this could be improved or worsen by changes to the overall budget of governments.

There are two broad and complementary ways in which public policies could facilitate reaching the SDGs: (1) modifying the opportunities and incentives of individuals and firms through micro-policies to, then, produce changes in the indicators; and (2) varying macro-policies through the selection of the budget size and the prioritization of government-funded programs that may affect the evolution of the indicators. In this paper, we analyze the second variant of policies and their potential for catalyzing improvements in the SDG indicators. Because the budget is traditionally modified on an annual basis, its size and allocation can be interpreted as short-term measures.

[1] The SDR is elaborated by the Sustainable Development Solutions Network and the Bertelsmann Stiftung.

Micro-policies, in contrast, relate to the creation of new behaviors, technologies, and organizational schemes; so they take more time to consolidate and impact the indicators. In this paper, we deploy a model that is suitable for studying macro-development policies, given the structural factors associated with already established historical micro-policies. Consequently, this framework allows estimating the impact on the feasibility of the SDGs when the budget of the existing government programs is modified.

Understanding the feasibility of the SDGs

Previous studies have attempted to assess the feasibility of subsets of SDGs by looking at historical data trends [1, 2]. However, assessing feasibility requires accounting for constraints on the growth of the indicators, and for their interdependencies. Such modeling conveys subtleties that differ from plain trend-based predictions of isolated indicators, for example, (i) while many societal actors contribute to the improvement of nations, development is unlikely to occur without properly funded government programs; (ii) public spending does not necessarily translate into development because there exist inefficiencies associated to poor governance, corruption, unnecessary bureaucracies, operative errors, shirking behavior, or incompetency; and (iii) improvements in certain policy issues may be confounded with improvements in other areas because of spillover effects coming from the presence of SDG interlinkages.

Simulating the dynamics of development indicators, while considering the aforementioned issues, requires methods that go beyond the analysis of data [3, 4, 5, 6, 7] and, instead, try to model behavioral and institutional data-generating mechanisms. A framework adhering to these premises can be found in the research program of Policy Priority Inference (PPI) [8, 9], which describes a policymaking process to simulate government decisions on the allocation of public expenditure and the indicator dynamics arising from these decisions.

This paper presents a new version of PPI that is able to account for different budget sizes and to measure how they affect the convergence towards a specific set of development goals. With this tool, we assess the following questions: (i) what will be the size of the remaining gaps between indicators and goals in 2030?; (ii) how sensitive is the evolution of the indicators to budgetary increments and reductions?; (iii) which is the maximum potential gap closure that could be achieved through sheer increases in government expenditure? Altogether, our study provides a comprehensive quantitative analysis of the feasibility of the SDGs for a large set of countries and indicators.

Data

We use data from the 2020 Sustainable Development Report (SDR), which contains time series on 79 indicators across 166 countries between the years 2000 and 2020.[2] In contrast to other SDG datasets, the SDR tries to minimize the number of missing observations and to avoid redundancies due to the use of disaggregate versions of the same indicator. Besides, it provides specific values for the goals to be reached by 2030 (a unique feature of this dataset). For this paper, we sample 140 countries based on the availability of complementary datasets such as aggregate government expenditure. We employ the Multi-Output Gaussian Process Toolkit [10] to impute missing observations[3] and, then, normalize the indicators between 0 and 1. Yet, if a nation has less than three observations for an indicator, the indicator is removed for this specific country. Thus, in our analysis, countries may have different indicators.

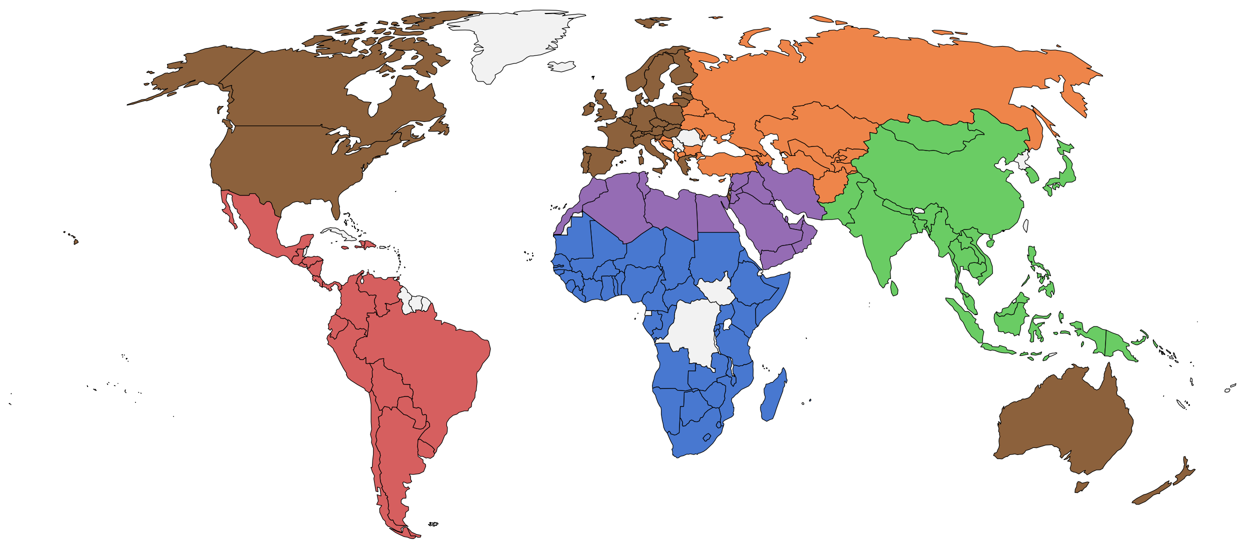
Although we perform all the estimations at the country level, we cluster the results into six groups. The clusters are Sub-Saharan Africa (*Africa*), Eastern Europe and Central Asia (*E. Europe & C. Asia*), East and South Asia (*East & South Asia*), Latin-America and the Caribbean (*LAC*), Middle-East and North Africa (*MENA*), and Western Countries (*West*). Figure 1 provides a map of the country groups covered in our sample.

We combine the data from the SDR with information from the following sources: the World Bank National Accounts Data and the OECD National Accounts data files, from which we obtain government expenditure information (the total expense and the net acquisition of nonfinancial assets); the Worldwide Governance Indicators [11], that presents information for *control of corruption* and *rule of law* which are helpful to empirically establish the quality of monitoring and law enforcement; the CEPII GeoDist Database [12] and the CEPII TradeProd dataset [13], from which we

[2] Due to the natural lags in reporting data, most of the missing values in the SDR dataset correspond to 2020, so this is where most extrapolations are performed. Nevertheless, there are some indicators that have observations for 2020. Importantly, the SDR acknowledges that these indicators are pre-COVID. Therefore, the estimates in our study may present optimistic scenarios.

[3] This method exploits information about the same indicator in a set of 'similar' countries to perform the imputations. Appendix B in the supplementary materials provides full details on how sets of similar countries are constructed and how the data is pre-processed.

Figure 1: Sampled countries and their clusters



Blue: *Africa*. Orange: *E. Europe & C. Asia*. Green: *East & South Asia*. Red: *LAC*. Purple: *MENA*. Brown: *West*. Countries in gray were excluded from the sample due to lack of data.

employ geographical and trade data for the data-imputation procedure. The interlinkages between indicators are estimated via the Sparse Gaussian Bayesian Networks approach developed by [14] using the time series from the indicators. All these data (including the networks) are specific to each nation, so our estimates consider context specificity in terms of the initial level of development, network topology, expenditure capacity, public governance effectiveness, and structural factors that account for the dynamics of the indicator that are not explained by PPI's model.

Results

Feasibility and expected gaps

Our estimations suggest that the SDGs Agenda is unfeasible if the aim were to close the gaps between the goals and indicators by 2030. Figure 2 shows that countries in *West* (the most advanced) still exhibit gaps of 12% on average. The estimated percentages are larger for the other five clusters. The worst performance corresponds to *Africa*, where the average gap is of 30% approximately. The rest of the clusters are closer to the 30% gap than to the one of *West*.

Notice that there is large heterogeneity in the size of these gaps across indicators, irrespective of the cluster. A common feature across all clusters is the presence of large gaps in SDG 9 ('Industry, Innovation, and Infrastructure').[4] This is consistent with [2], who use different indicators and find poor historical performance in SDG 9.

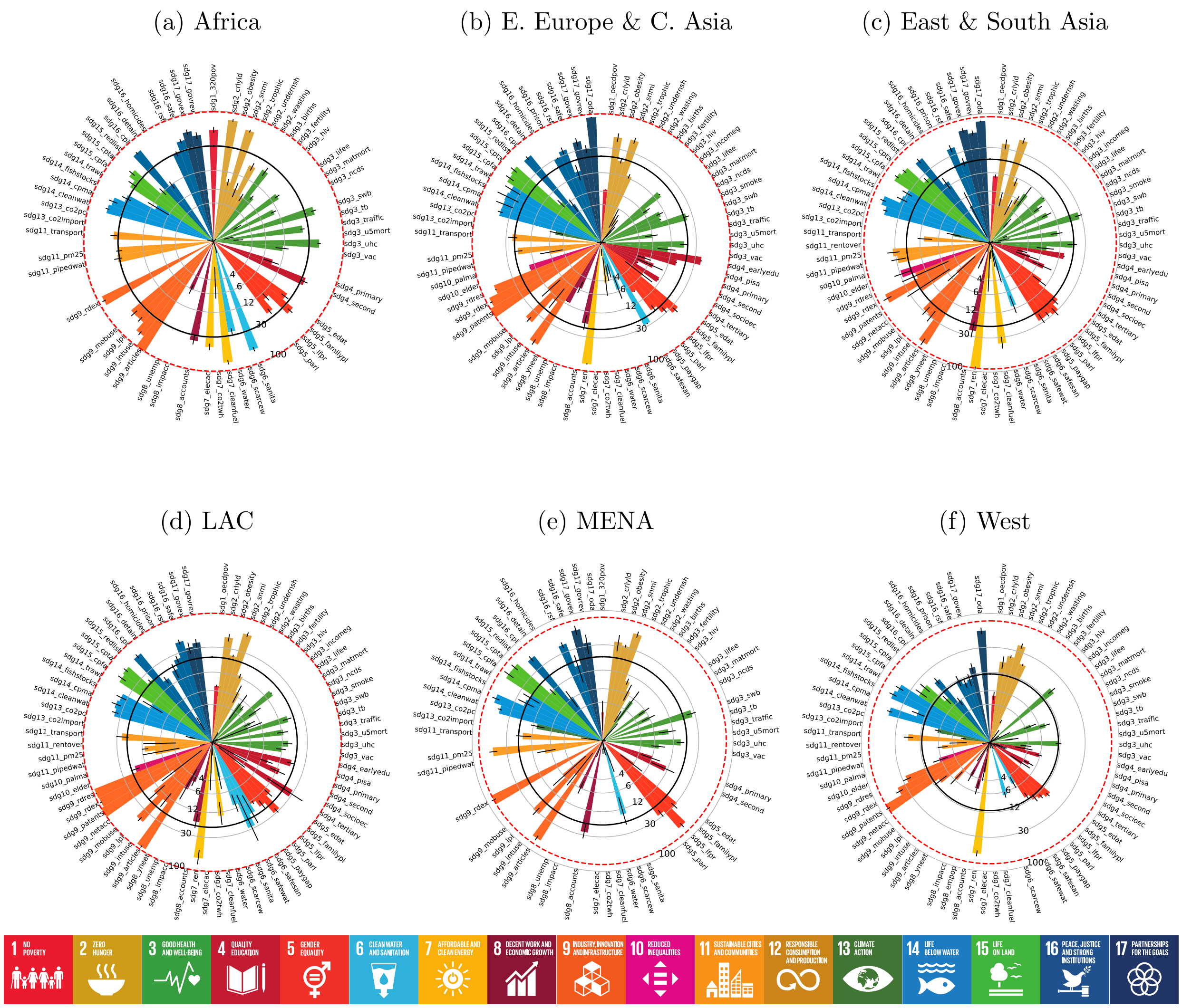
Most policy issues in SDGs 14 and 15 show gaps above the average for all the clusters. However, this visualization also highlights discrepancies between clusters when looking at the same indicators. For instance, three indicators of SDG 16 ('Peace, Justice and Strong Institutions'), present gaps above their cluster average in *LAC*, while there is only one gap above the average in *West* for this SDG.

In Figure 3, we show that the gaps, averaged across indicators, are non-negligible across all the countries included in the sample, whether they have developing, emerging, or advanced economies. Nonetheless, we detect substantial cluster variation, as well as discrepancies between countries located in the same group. For example, in *East & South Asia*, Japan (JPN) presents an average gap for the year 2030 (height of the faded bar) close to 12%. In contrast, such a gap is larger than 30% in Pakistan (PAK). Consistent with the results presented in Figure 2, *Africa* exhibits the worst performance and *West* has the best one. However, several countries outside *West* show similar gaps to those prevailing in advanced nations, for example, Bulgaria (BGR), Belarus (BLR), Cyprus (CYP), Croatia (HRV), Japan (JPN), South Korea (KOR), Singapore (SGP), and the United Arab Emirates (ARE).

Would the gaps shrink proportionally if the SDGs were extended an extra 20 years? To answer this, we estimate the gaps expected by 2050, and show them in Figure 2 as the solid bars. Notice that the size of the gap reduction is relatively limited, even though the model runs for an additional 20 years. This implies the existence of decreasing returns in the continued usage of budgetary policies to attain the SDGs. Then, can these decreasing returns be compensated by increasing sheer spending? To answer this question, we analyze the sensitivity of the indicator dynamics to budgetary changes.

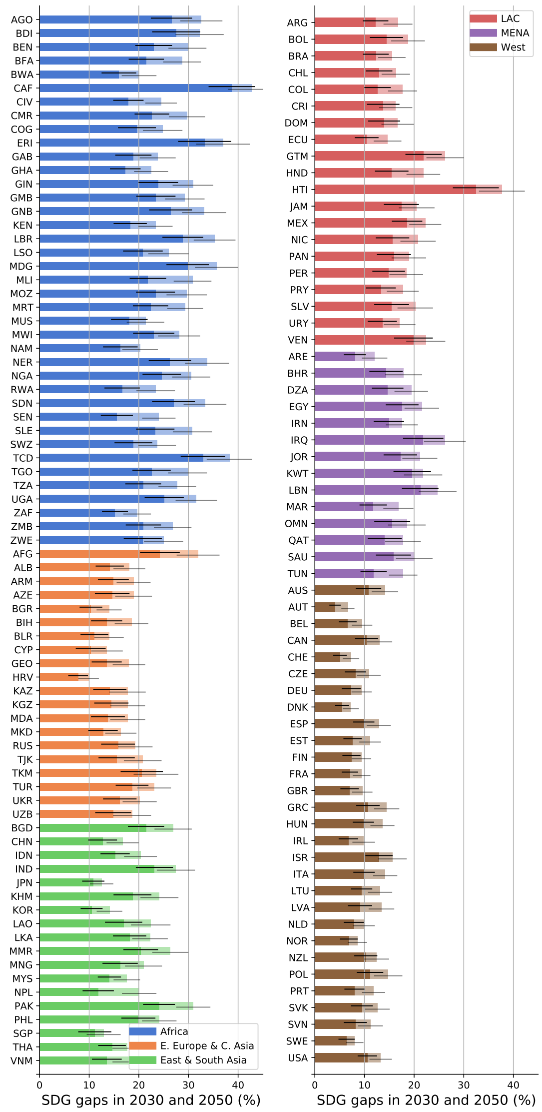
[4] The full names of these indicators are presented in Table B.3 of the supplementary materials.

Figure 2: Average gap by 2030



Height of each bar: average gap between the SDG and the indicator level predicted by 2030 computed across countries in the cluster. Empty spaces between bars: no data was available for the corresponding indicator in any country from the cluster. Solid black ring: average gap across across countries (in the cluster) and indicators. Dashed red ring: largest average gap (between indicators in the cluster). Black lines at the top of each bar: ± standard error of the mean gaps across the countries of a cluster.

Figure 3: Average gaps by country



Faded bars: average gaps (%) between SDGs and the indicator levels predicted by 2030. Solid bars: average gaps (%) between SDGs and the indicator levels predicted by 2050. Grey lines: ± standard error of the 2030 mean gaps across the indicators of each country. Black lines: ± standard error of the 2050 mean gaps across the indicators of each country.

Sensitivity to changes in the budget size

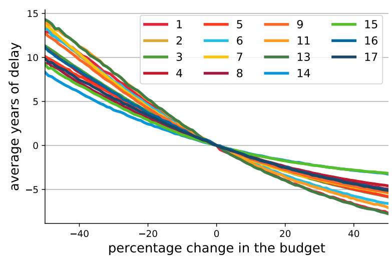
In order to produce sensitivity estimates, we assume that the budget for the 2020-30 decade is equivalent to 10 years of public spending according to the historical levels shown by each country in the dataset (in real terms). Of course, the size of a budget may change as governments see it fit, for example, inflation-adjusted government expenditure can grow over time due to unexpected events or to expansionary policies. Nevertheless, budgetary changes exhibit universal and stable distributions across different countries and periods [19]. Our dataset suggest substantial variation in the growth of public spending between the 2000-10 and the 2010-20 decades (an average of 47%). However, a growth scenario for the next ten years may be hindered thanks to the COVID-19 global pandemic. In general, the final effects of external shocks and government interventions in a 10-year horizon cannot be known with certainty, but it is possible to discover potential outcomes by simulating scenarios in which countries experience reductions (or increments) to their overall budgets. These counterfactuals allow us to understand the sensitivity to positive and negative changes in the size of public spending.

Three important assumptions about our modeling approach should be highlighted to interpret our results. First, PPI aims to model short-term dynamics and, hence, long-term structural factors are given through the exogenous parameters α_i that are specific to each indicator and country. Second, the impact of the public funds devoted to the different government programs is viewed in the context of short-term effects. This is so because we model a probability $\gamma_{i,t}$ representing the chance of indicator i to improve in the subsequent period $t+1$. These two aspects are combined into the law of motion of indicator i: $I_{i,t+1} = I_{i,t} + \alpha_i \xi(\gamma_{i,t})$, where ξ is the binary outcome of a successful policy implementation with probability $\gamma_{i,t}$ (which is a random event). While more public spending increases this probability, the long-term structural factors α_i limit the speed of such growth. Therefore, government expenditure only affects $\gamma_{i,t}$, not α_i. Third, public spending is a necessary condition for development. It is clear from the law of motion that, if less expenditure brings $\gamma_{t,i}$ close to zero, then the growth trials are almost-always unsuccessful, so the indicator dynamics stagnate.

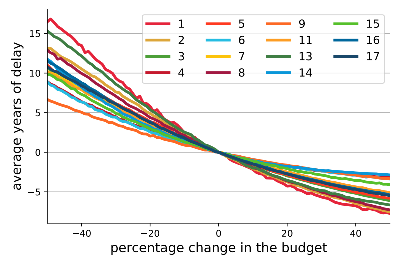
In Figure 4, we quantify the average increment(reduction) in the number of years required to close the SDG gaps. These results are produced by generating a series of 1% changes in the budget of a given country and running the model for an equivalent of 50 years. The range of budgetary changes goes from -50% to +50%. The estimations are performed for each country and budgetary change individually (each one requiring 1000 Monte Carlo simulations), but the plots present the aggregated curves for each cluster.[5]

Figure 4: Changes in convergence time as a function of the budget size

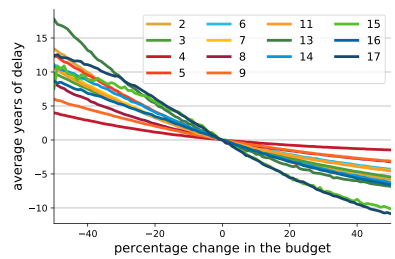
(a) Africa



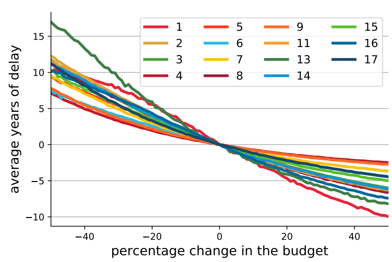
(b) E. Europe & C. Asia



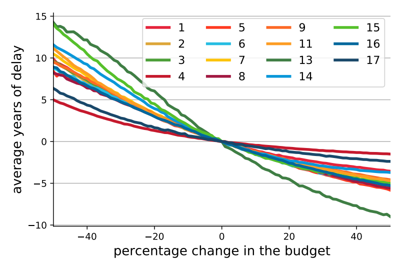
(c) East & South Asia



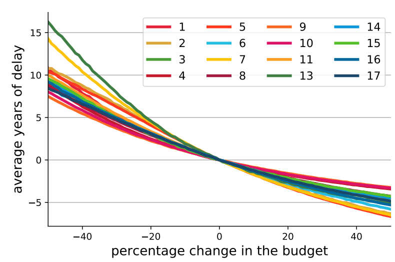
(d) LAC



(e) MENA



(f) West



First, notice that every indicator shows certain level of sensitivity to both positive and negative budgetary changes. For instance, indicators in SDG 1 ('No Poverty') in *LAC* exhibit time savings of 10 years when the overall budget increases by 50% while, in the same cluster, indicators in SDG 13 ('Climate Action') show a delay of more than 15 years when the budget shrinks 50%. Second, the sensitivity to budgetary changes are asymmetrical in terms of the responses' magnitudes. While delays fluctuate between 4 and 17 years for a 50% reduction in the budget size, savings in conver-

[5]The curves in Figure 4 correspond to indicators that, in the -50% scenario, were able to converge in 50 years or less. Convergence among indicators with an extremely low α_i may be too slow, making this exercise implausible for all countries and budgetary changes considered. Instead, since the figure intends to illustrate aggregate results, we run the equivalent of 50 years and discard those indicators that did not close the gaps in that time.

gence time fluctuate between 1 and 10 years for a 50% budgetary increment. By construction of the model, such asymmetry is expected since convergence time tends to infinity as $\gamma_{i,t}$ approaches zero. However, its presence under empirically plausible budgetary changes (not only in theoretical limits) speaks to the importance of linking public spending and indicator dynamics.

Importantly, the sensitivity ranking across indicators varies between clusters, magnitude, and direction of the budgetary change. For example, for countries in *West*, SDG 13 is the most sensitive to budgetary reductions. However, its rank as the most sensitive SDG is overtaken by SDG 5 when the budget grows. Notice that, in all clusters, SDG 13 exhibits important delays when there are budget cuts. It means that environmental variables such as *clean air* systematically under-perform more than others when public spending shrinks.

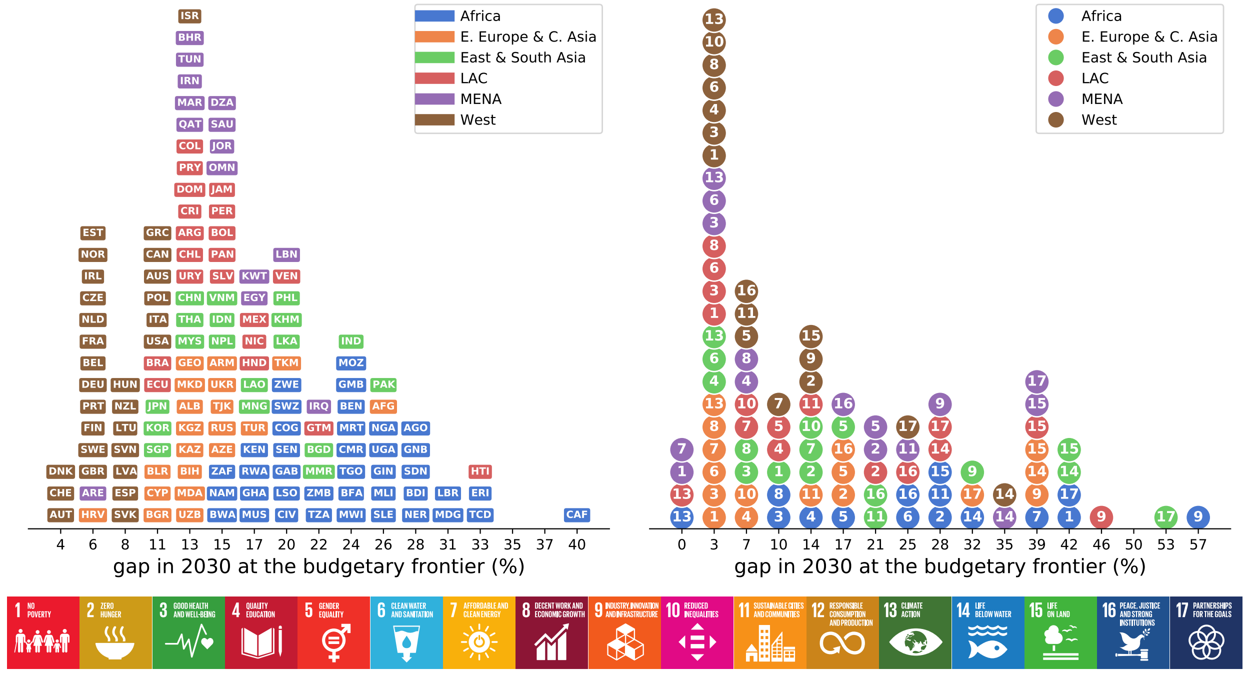
Budgetary frontiers and structural bottlenecks

Long-term structural factors constrain the capacity that governments have in the short-run for accelerating development through sheer spending. To quantify these structural bottlenecks, we introduce the concept of *budgetary frontier*. A budgetary frontier is a–hypothetical–situation in which countries develop without being limited by the amount and efficiency of public spending. Formally, this translates into setting $\gamma_{i,t} = 1$ in the model. This concept assumes that public policies have no financial constraints and that their impact is effective at all time. In this hypothetical setting, development trajectories are limited only by the long-term structural factors α_i. The budgetary frontier is useful to identify and compare the different structural challenges prevailing across countries, clusters, and SDGs; i.e., challenges that require more than budgetary increments.

We report the outcome of simulations in which countries operate at the budgetary frontier. In this exercise, we estimate the average gap that would remain in 2030 if countries would operate at the budgetary frontier. The left panel in Figure 5 presents the average gaps across indicators, at the country level. The right panel shows the average gaps across indicators and countries, at the SDG-cluster level. Note that, in the left panel, the gaps do not close entirely, even in countries located in the *West* cluster. This suggests the existence of structural bottlenecks that would have to be overcome if the 2030 Agenda were to be reached. As expected, the gaps at the budgetary frontiers are the largest among countries in *Africa*, being as high as 40% for the Central African Republic (CAF).

Figure 5: Budgetary frontiers

(a) Country level (b) SDG-cluster level



A country operating at the budgetary frontier has a $\gamma_{i,t} = 1$ for every indicator i and every period t (see equation 2) in the methods section). At the budgetary frontier, the only frictions slowing down the indicators' growth are the structural parameters α_i. Panel a: budgetary frontiers calculated by averaging gaps across indicators for each individual country. Panel b: budgetary frontiers calculated by averaging gaps across countries and indicators at the level of SDG and country-cluster combinations. The average gaps have been discretized in order to produce the visualizations.

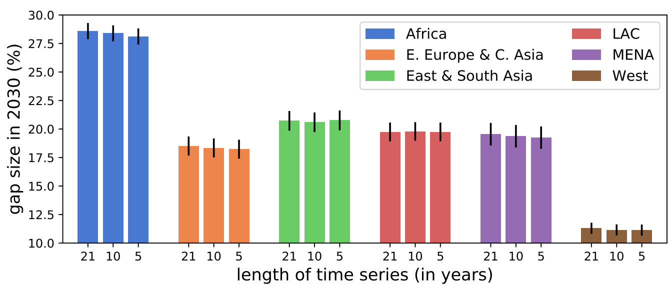
The right panel in Figure 5 shows that the average gaps at the budgetary frontier, for the different SDGs and clusters, vary between 0 and 57%. From this, we gather that the *West* cluster's SGDs concentrate in the lower spectrum of the budgetary frontiers, signaling relatively good performance. In contrast, the corresponding gaps for *Africa* are dispersed throughout the spectrum, finding the largest bottleneck in SDG 9. The fact that SDG 13 presents a near-null gap for countries in the *Africa* and *LAC* clusters indicates that environmental issues related to climate action could be remedied, in the short-term, by properly funding existing policies. However, other SDGs related to sustainability show worse performances. This is the case of SDGs 14 and 15, whose estimates range between 28 and 42%[6], and whose worst performers are in *East & South Asia*.

Finally, the reader may consider that public spending should have structural consequences, so

[6] SDG 15 for *West* is an outlier with a gap of 14%.

the exogenous factors α_i calibrated with the empirical data could also be affected in the short term. The empirical evidence suggests otherwise. If the structural factors contained in α_i would significantly change in the short-term, then the estimated gaps for simulations that use more recent data samples should significantly differ. To demonstrate that this is not the case, we calibrate and perform the same analysis but, instead of using the full 21-year dataset (with 2000-2020 coverage), we employ a 10-year (2011-2020) sample and a 5-year (2016-2020) one.[7] Figure 6 shows that our estimates are robust to these samples, as the six clusters show practically no difference in their average gaps (see Appendix G for more disaggregate yet robust results). From this, we conclude that the SDG network and the structural factors exhibit slow dynamics, validating our conceptualization of long- versus short-term effects. Accordingly, the budgetary frontiers involve long-term considerations and demand the implementation of innovative micro-policies.

Figure 6: Robustness to different sampling lengths



Each bar represents the average gap estimated for 2030 across indicators and countries within each cluster. The vertical black lines correspond to the ± standard error.

Limitations and conclusion

PPI is a framework designed to perform short- and mid-term analysis. For this, it assumes fixed structural factors for the indicators' growth and fixed network topologies for the interlinkages between indicators. It is not designed to identify the causes behind the structural constraints that prevent a country from closing its SDG gaps. Hence, PPI's outputs are not informative about how to reformulate the existing micro-policies or how to generate new ones. Instead, PPI is useful to

[7]This involves re-estimating the network, the structural parameters, and the gaps.

understand how to allocate resources across the existing micro-policies. Likewise, its results have to be interpreted with caution when micro-policies change significantly with respect to those that existed in the past. However, through a counter-factual exercise, PPI's underlying computational model can identify the bottlenecks associated with the inefficacy of the budget in fostering specific issues.

The main results of our simulations indicate that, in general, the established goals for 2030–according to the Sustainable Development Report–are unfeasible even for advanced countries. We find that the indicators are, in general, sensitive to the magnitude and direction of changes in the overall government budget, although in a non-linear and asymmetric fashion. Our results suggest that more government spending is not enough to close the SDGs gaps, even if countries operate at the budgetary frontier. Hence, complementary micro-policies are ultimately needed to overcome structural–long-term–bottlenecks and to strengthen the dynamics of the indicators. Finally some of the environmental concerns (e.g., clean air) can be remedied with a larger budget while others (e.g., SDG 14 and 15) require the undertaking of well-designed government programs for shifting historical inertias.

Methods

Model

There are N policy issues, and the government can directly intervene $n \leq N$ of them through public policies. Using the PPI jargon, we say that there are n instrumental issues and $N - n$ collateral ones. An issue may be collateral because there does not exist a policy instrument to directly intervene it, and this may occur because the issue is too aggregate (e.g., GDP growth). The level of development of each issue is quantified through an indicator. For each indicator, the government has a goal that it wants to achieve. Thus, there is a vector $T_1, \ldots, T_N$ of development goals, which is an exogenous variable (in our empirical application, such vector is obtained from the SDR dataset).

The model contains two types of agents: one central authority and n policymaking agencies (one per instrumental issue). The central authority is responsible for allocating resources across the n instrumental issues in order to achieve its goals. The policymaking agents, however, may only use a fraction of the allocated resources since their incentives are not perfectly aligned with those of the central authority. To be more precise, for a given period t, let $P_{i,t}$ denote the amount of resources that the central authority has decided to allocate to policy issue i, and $C_{i,t} \leq P_{i,t}$ is the fraction of resources that are effectively used by the corresponding agent to improve indicator i. Then, we say that $P_{i,t} - C_{i,t}$ is the level of inefficiency in policy issue i.

Variables $P_{i,t}$ and $C_{i,t}$ are dynamic and endogenous. On the one hand, the central authority (or government) determines $P_{i,t}$ through an adaptive heuristic of policy priorities while, on the other, the policymaking agents (or public officials) establish $C_{i,t}$ via a behavioral rule of reinforcement learning.[8] The government's expenditure heuristic prioritizes those issues that lag behind the most[9] and those that show to be historically more efficient. The central authority gathers information on historical efficiencies through its monitoring mechanisms, which may vary in quality between countries. More specifically, there is a stochastic process for spotting inefficiencies, and its probability of success depends on the quality of monitoring[10] as well as on the size of the inefficiency $P_{i,t} - C_{i,t}$. When public officials are caught spending less than $P_{i,t}$, they are penalized according to the quality of the rule of law in the country.[11] This penalty feeds back into their experience and contributes to pondering whether to increase or decrease $C_{i,t+1}$. This learning process is motivated by the rewards the agents receive from the personal gain $P_{i,t} - C_{i,t}$ (minus the penalty if detected) and the political status obtained from improving the associated indicators.

Regarding the indicators, let us define their law of motion according to

$$I_{i,t+1} = I_{i,t} + \alpha_i \xi(\gamma_{i,t}), \tag{1}$$

where parameter $\alpha_i \in (0, 1)$ is exogenous and requires calibration. In [8], α_i is referred to as the growth factor of the indicator, and it is interpreted as a residual accounting for all other determinants responsible for an indicator's growth that are not explicit in the model. However, a more workable interpretation of α_i is that it captures mainly structural factors. This interpretation is useful when considering public expenditure data because α_i contributes to imposing a limit to the growth that could be achieved in the short-term by sheer spending.

[8] A more detailed exposition of the model is presented in Appendix A.
[9] A promoted practice during the Millennium Development Project.
[10] An exogenous parameter coming from public governance data.
[11] Another exogenous parameter coming from public governance data.

To see this more clearly, let $\xi(\gamma_{i,t})$ in equation 1 denote the outcome of a Bernoulli trial that can take values 1 (successful) or 0 (unsuccessful). This means that, if a positive event materializes, the indicator grows according to α_i. The probability of a successful trial is $\gamma_{i,t}$, which is one of the dynamic and endogenous variables of the model. It is through this channel that public spending exerts its influence on the indicator dynamic.

An implicit assumption in [8] is the conceptualization of the budget as a flow variable, not as a stock one. In order to infer the impact of budget size on SDGs feasibility, we need to reformulate this concept to allow for a stock perspective (i.e., the accumulated disbursements during the period of analysis). Let us assume that the government has a total budget of size $\boldsymbol{B}$ to be spent over a certain period. This stock can be turned into flows by defining a disbursement schedule $B_1, \ldots, B_{\mathcal{T}}$, such that $\sum_t^{\mathcal{T}} B_t = \boldsymbol{B}$. This schedule implies that $\boldsymbol{B}$ is depleted by time $\mathcal{T}$ (unless the stock is re-filled). For simplicity, let us assume that the disbursement schedule is homogeneous, so $B_t = B \quad \forall \quad t$.

Consider the allocation profile $P_{1,t}, \ldots, P_{n,t}$ that the central authority defines in time t. Under the homogeneous disbursement schedule assumption, $\sum_i^n P_{i,t} = B$ holds.[12] When budgetary data is provided to the model, variables $P_{i,t}$ and $C_{i,t}$ are defined in monetary units. In order to map the monetary variable $C_{i,t}$–reflecting the public officials' contributions–into the success probability $\gamma_{i,t}$, we define

$$\gamma_{i,t} = \beta \frac{C_{i,t} + \frac{1}{n}\sum_j C_{j,t}}{1 + e^{-S_{i,t}}}, \tag{2}$$

where β is a normalizing parameter and $S_{i,t}$ is the net amount of spillovers received by indicator i in period t (this could be positive or negative). The spillovers are computed every simulation period according to $S_{i,t} = \sum_j \mathbf{1}_{j,t} \mathbb{A}_{j,i}$, where $\mathbf{1}$ is an indicator function that returns 1 if indicator j grew in the previous period and 0 otherwise. The adjacency matrix $\mathbb{A}$ corresponds to the network of interlinkages, with each entry representing a conditional dependence calculated according to the method established by [14]. Importantly, these conditional dependencies do not represent causal links, but rather an empirical regularity that PPI takes into account (see [15] for a detailed discussion on estimating SDG networks).

This formulation is an improvement over [8] in several ways. First, parameter β helps translating

[12] In [8], it is assumed that $B = 1$ and there is no notion of a stock budgetary variable since $\mathcal{T}$ is calibrated. Under this new version of PPI, B can take monetary units, and $\mathcal{T}$ defines the frequency of the disbursement schedule.

the monetary units from $C_{i,t}$ and $\bar{C}_t$ into a probability. Hence, for a given β, equation 2 is sensitive to changes in $C_{i,t}$ and, consequently, to the budget size. Second, we introduce $\frac{1}{n}\sum_j C_{j,t}$ in order to capture the general 'health' of public finances; i.e., the overall size of the budget effectively used to improve the indicators' performance. Third, in contrast to [8], α_i appears only in equation 1 (previously, it was also part of equation 2). This helps separating the short-term effects of public spending from long-term structural transformations (which could be described in the model in terms of changes in α_i).

Calibration procedure

Because the context of each nation is important, the model is calibrated independently for each country. This procedure consists of fitting the free parameters such that the simulated indicators capture specific features of the empirical ones. Simulations are performed for the sampling period 2000-2020, where the initial values of the indicators correspond to the ones in 2000, and the vector $T_i, \ldots, T_N$ is determined by the values observed in 2020. In PPI these are called the retrospective simulations. Thus the calibration seeks to tune the model so that it explains why the historical indicators moved from their initial values and arrived to the final ones in the last observation period. The calibration also seeks to account for the volatility of the indicators in terms of the frequency with which positive changes are observed.

There are two types of parameters that need calibration: the growth factors $\alpha_1, \ldots, \alpha_N$ and the normalizing constant β. To calibrate the growth factors, we improve the fitting algorithm provided by [8]. The objective function seeks to minimize the difference between the empirical indicator in 2020 and its synthetic counterpart in the last simulation period by choosing an adequate α_i. The final simulated data point comes from the average final value of M Monte Carlo simulations. To find the full vector $\alpha_1, \ldots, \alpha_N$, the order of the indicators is randomized. Then, a bounded optimization procedure is performed for indicator i using Brent's method (the bounds are 0 and 10). Once the optimal α_i is found, the vector of structural factors is updated and the algorithm moves on to the next indicator.

To calibrate β, we develop a straightforward procedure. Since this parameter has a direct incidence in the probability of positive growth, we want to find a β such that $\gamma_{i,t}$ is close to the empirical quantity $\hat{\gamma}$, which is the fraction of positive changes observed across all indicators and years in the data. We employ Brent' method to find β. Each evaluation of β computes the average γ_i across indicators, simulation periods, and Monte Carlo simulations. We, then, combine these two protocols in the iterative calibration process described in algorithm 1.

The proposed algorithm performs the optimization of β first and, then, the one of vector $\alpha_1, \ldots, \alpha_N$.[13] These two procedures are repeated until an error threshold is passed. The error of β is the absolute value of the difference between the empirical success rate and the average γ across indicators, simulation periods, and Monte Carlo simulations. Its corresponding threshold is set at 0.0075. For a single indicator i, its error is the difference (in absolute value) between its most recent empirical value and the average final value obtained from Monte Carlo simulations performed for a given α_i. Then, the error of vector $\alpha_1, \ldots, \alpha_N$ is the average of the individual errors of the indicators. We establish a threshold of 0.0025 for this type of error. If the thresholds are not met after 10 iterations, we use the best-performing solution. Overall, the algorithms usually meets the error thresholds (see appendix E in the supplementary materials for the reported errors). These thresholds could be more stringent, but the procedure would require more simulations per evaluation, which becomes computationally expensive for 140 countries. From performing the same analysis on a reduced sample, we find that the results are robust for different threshold levels.

Algorithm 1: Calibration procedure

Input: empirical data

while $\beta_{error} > \beta_{tolerance}$ *or* $\alpha_{error} > \alpha_{tolerance}$ **do**
- optimize β;
- randomize order of indicators;
- **foreach** *indicator i* **do**
  - find optimal α_i;
- update β_{error} and α_{error};

Estimation of the SDG gaps

To simulate the indicators evolving from 2020 onward, we use the parameters β and $\alpha_1, \ldots, \alpha_N$ previously calibrated for each country. The goals $T_1, \ldots, T_N$ are provided by the dataset of the SDR. Each simulation is run for the equivalent of 10 years. For each indicator, we measure the

[13] This execution order is motivated by the fact that [8] show little sensitivity in the indicators' volatility to changes in α.

distance between its final value and T_i. If the indicator surpasses the goal, then the gap is set at zero.[14] This computation is performed for each Monte Carlo simulation, so we take the average gap (distance) across simulations. When presenting aggregate results, we average these quantities across indicators and countries in the same cluster.

A set of Monte Carlo simulations consists of 1000 independent simulations. For the calibration part, we set a disbursement schedule of 50 simulation periods throughout the 21 years of data. For the prospective part, we run the model for the number of simulation periods corresponding to 10 years and, then, we calculate a gap for each indicator. Appendix F shows that varying the number of simulation periods for the same dataset yields robust results. All other results in the paper come from running counterfactual simulations and measuring the corresponding gaps. For sensitivity to the size for the budget, the counterfactual consists of changing the size of the government expenditure datum in increments (and decrements) of 1%. Then, for the budgetary frontiers, the probability of success γ_i is forced to be 1, having the same effect as a 100% efficient and non-constrained budget.

References

[1] González-Pier, E., Barraza-Lloréns, M., Beyeler, N., Jamison, D., Knaul, F., Lozano, R., Yamey, G., & Sepúlveda, J. Mexico's path towards the Sustainable Development Goal for health: An assessment of the feasibility of reducing premature mortality by 40% by 2030. The Lancet Global Health, 4(10), e714–e725 (2016).

[2] Luken, R., Mörec, U., & Meinert, T. Data quality and feasibility issues with industry-related Sustainable Development Goal targets for Sub-Saharan African countries. Sustainable Development, 28(1), 91–100 (2020).

[3] Fuso-Nerini, F., Sovacool, B., Hughes, N., Cozzi, L., Cosgrave, E., Howells, M., Tavoni, M., Tomei, J., Zerriffi, H., & Milligan, B. Connecting climate action with other Sustainable Development Goals. Nature Sustainability, 2, 674–680 (2019).

[14] If the goal is the same as the theoretical maximum of the indicator, the model prevents the indicator from growing further.

[4] Lusseau, D. & Manchini, F. Income-based variation in Sustainable Development Goal interaction networks. Nature Sustainability 2, 242–247 (2019).

[5] McGowan, P., Stewart, G. Long, G. & Grainger, M. An imperfect vision of indivisibility in the Sustainable Development Goals. Nature Sustainability volume 2, 43–45 (2019).

[6] Kroll, C., Warchold, A. & Pradhan, P. Sustainable Development Goals (SDGs): Are we successful in turning trade-offs into synergies?. Palgrave Communications volume 5, 140 (2019).

[7] Porciello, J., Maryia, I., Islam, M., Einarson, S., & Hirsh, H. Accelerating evidence-informed decision-making for the Sustainable Development Goals using machine learning. Nature Machine Intelligence volume 2, 559–565 (2020).

[8] Guerrero, O. & Castañeda, G. Policy Priority Inference: A Computational Framework to Analyze the Allocation of Resources for the Sustainable Development Goals. Data & Policy forthcoming, (2020).

[9] Castañeda, G., Chávez-Juárez, F. & Guerrero, O. How Do Governments Determine Policy Priorities? Studying Development Strategies through Networked Spillovers. Journal of Economic Behavior & Organization 154, 335–361 (2018).

[10] de Wolff, T., Cuevas, A. & Tobar, F. MOGPTK: The Multi-Output Gaussian Process Toolkit. arXiv:2002.03471 (2020).

[11] Kaufmann, D., Kraay, A. & Mastruzzi, M. The Worldwide Governance Indicators: Methodology and Analytical Issues. Hague Journal on the Rule of Law 3, 220–246 (2011).

[12] Mayer, T. & Zignago, S. Notes on CEPII's distances measures: The GeoDist database. (2011).

[13] de Sousa J., Mayer, T. & Zignago, S. Market Access in Global and Regional Trade. Regional Science and Urban Economics. 42(6), 1037-1052 (2012).

[14] Aragam, B., Gu, J. & Zhou, Q. Learning Large-Scale Bayesian Networks with the Sparsebn Package. Journal of Statistical Software 91, (2019).

[15] Ospina-Forero, L., Castañeda Ramos, G., & Guerrero, O. Estimating Networks of Sustainable Development Goals. JInformation and Management, (2020).

[16] Lafortune, G., Fuller, G., Moreno, J., Schmidt-Traub, G. & Kroll, C. SDG Index and Dashboards. Detailed Methodological Paper. (2018).

[17] Sachs, J. et al. Sustainable Development Report 2020. (Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN), 2020).

[18] World Bank & OECD. World Bank National Accounts Data, and OECD National Accounts Data Files. (2020).

[19] Jones, B., Baumgartner, F., Breunig, C., Wlezien, C., Soroka, S., Foucault, M., François, A., Green-Pedersen, C., Koski, C., John, P., Mortensen, P., Varone, F., & Walgrave, S. A General Empirical Law of Public Budgets: A Comparative Analysiss. American Journal of Political Science, 53(4), 855–873. (2009).

Data availability statement

All raw data are publicly available through the original sources. All processed data, calibrated parameters, and the simulation outputs are available in the repository `github.com/oguerrer/SDG_feasibility`.

Code availability statement

The code with the mew version of PPI and for its calibration is available in the repository `github.com/oguerrer/SDG_feasibility`.

How Feasible are the Sustainable Development Goals?

Analyzing Government Spending and Policy Priorities Worldwide

Supplementary Materials

Omar A. Guerrero[1,2] and Gonzalo Castañeda[3]

[1]Department of Economics, UCL, United Kingdom

[2]The Alan Turing Institute, United Kingdom

[3]Centro de Investigación y Docencia Económica (CIDE), Mexico

A Model details

The Policy Priority Framework (PPI) is explained in great detail in [8]. However, since we have made important improvements to its original formulation, it is useful to provide a self-contained exposition of all the equations in the model. Nevertheless, our explanation will spare the reader from interpretations, motivations, and references to the literature, as they remain close to the original ones.

A.1 Policy-making agents

There are n agents (or public officials), each in charge of a public policy that is specific to a single policy issue. To implement the mandated policy in a given period t, agent i receives $P_{i,t}$ resources from the central authority (or government). With these resources, the public official tries to leverage two potential benefits: (1) the reputation from being a proficient public servant and (2) the utility derived from being inefficient according to

$$F_{i,t+1} = \Delta I^*_{i,t} \frac{C_{i,t}}{P_{i,t}} + (1 - \theta_{i,t}\tau) \frac{(P_{i,t} - C_{i,t})}{P_{i,t}}, \tag{3}$$

where $F_{i,t+1}$ represents the benefit or utility obtained in the next period. The first summand in equation 3 captures the benefit of being proficient. $\Delta I^*_{i,t}$ is the change in indicator i with respect to the previous period (its performance), relative to the changes of all other indicators. More specifically, the relative change in indicator i is computed as

$$\Delta I^*_{i,t} = \frac{I_{i,t} - I_{i,t-1}}{\sum_j I_{j,t} - I_{j,t-1}}, \tag{4}$$

and it captures the idea that the central authority compares and evaluates the relative performance of each public official, and their implemented policies, through the corresponding development indicators.

Going back to the first summand in equation 3, we find that the relative change in the indicator is pondered by $\frac{C_{i,t}}{P_{i,t}}$. Here, $C_{i,t}$ is the fraction of the allocated resources $P_{i,t}$ that are effectively used towards the policy. We call it the *contribution* of agent i.

Next, let us focus on the second addend of equation 3, which corresponds to the utility derived from being inefficient. Here, $P_{i,t} - C_{i,t}$ is the benefit extracted from not devoting resources to the policy. Thus, when dividing by $P_{i,t}$, it represents the level of inefficiency. Monitoring and penalties may hinder inefficiencies. This is captured by factor $(1-\theta_{i,t}\tau)$. Variable $\theta_{i,t}$ is the binary outcome of monitoring inefficiencies. If $\theta_{i,t} = 1$, it means that the government has spotted agent i in inefficient behavior. In that case, i is penalized by a factor τ, such that the benefit from these private gains are reduced.

In order to model the binary outcomes of monitoring efforts, we assume that, every period, an independent realization of $\theta_{i,t}$ takes place for each indicator. This is nothing else than a Bernoulli process with a probability of success $\lambda_{i,t}$ determined by

$$\lambda_{i,t} = \varphi \frac{P_{i,t} - C_{i,t}}{P^*_t}, \tag{5}$$

where P^*_t is the largest allocation in period t. Parameter φ in equation 5 corresponds to the quality of the monitoring efforts.

If an agent becomes more inefficient and their benefits increase, then reinforcement learning takes place, becoming more inefficient the next period. If, in contrast, the government is able to penalize, according to the learning process, they become more proficient the next period. Formally, action $X_{i,t}$ of agent i can be modeled as

$$X_{i,t+1} = X_{i,t} + \text{sgn}((X_{i,t} - X_{i,t-1})(F_{i,t} - F_{i,t-1}))|F_{i,t} - F_{i,t-1}|, \tag{6}$$

where $\text{sgn}(\cdot)$ is the sign function. In order to map action $X_{i,t}$ into the value of the effective resources, we define

$$C_{i,t} = \frac{P_{i,t}}{1 + e^{-X_{i,t}}}. \tag{7}$$

A.2 The government agent

Policy priorities are represented by the allocation profile $P = P_1, \ldots P_n$. It is important to introduce a distinction between those indicators that can be intervened via public policies: *instrumental*; and those that cannot: *collateral*. An instrumental indicator exists if the government has a policy or program to directly impact it (i.e., it receives resources). In contrast, a collateral indicator cannot be directly impacted; it is a composite aggregation of various topics, for example, GDP per capita or financial development. Policy priorities can only be defined on the n instrumental indicators, while there can only be n public officials (one in charge of each instrumental indicator). When talking about all the indicators together, we say that there are $N \geq n$ policy issues in total.

The objective of the government is to close the gap between the goals and the indicators by solving the problem

$$\min \left[\sum_{i}^{N} (T_i - I_{i,t})^2 \right] \tag{8}$$

through the allocation of budgetary resources across policy issues. The central authority achieves this by adapting its allocation profile P.

In the real world, identifying the precise mechanisms through which governments establish their budgets is extremely challenging. A starting point is the principle of 'gaping', which suggests that governments prioritize the most laggard topics as these may be development bottlenecks. Nevertheless, the political process also introduces adaptations motivated from signals such as the people's demands, and the performance of the different expenditure programs. In the political science literature, these budgetary changes exhibit punctuated dynamics and are modeled through simple stochastic processes [19]. Thus, we combine all these insights in a government heuristic where the policy priorities are established according to

$$P_{i,t} = B \frac{q_{i,t}}{\sum_j q_{j,t}}, \tag{9}$$

where $q_{i,t}$ is the propensity to spend in policy issue i in time t, and B is the budget available in time t.

The evolution of the policy priorities takes place through the propensities. In the first period, these are determined by the normalized gaps

$$q_{i,0} = \frac{T_i - I_{i,0}}{\max(T_{\cdot} - I_{\cdot,0})}. \tag{10}$$

Then, as time progresses, the propensities are updated according to

$$q_{i,t} = q_{i,t-1} + U(0,1) \left(\sum_{k}^{t-1} \theta_{i,k} \right)^{-1} \sum_{k|\theta_{i,k}=1}^{t-1} \frac{P_{i,k} - C_{i,k}}{P_{i,k}}. \tag{11}$$

The previous equation is rather intuitive. The term $U(0,1)$ is a random draw from a uniform distribution in the (0,1) interval. This captures the randomness of societal signals received by the government (it is consistent with the stochastic processes used to model budgetary changes in the literature). The remaining terms to the right correspond to the inter-temporal average inefficiency, which lies in the interval [0,1]. Therefore, the government encourages increments among the most efficient policymaking agents. Note that, in general, the contribution $C_{i,t}$ is not observable by the government, unless there is a successful audit by the monitoring authority. This is why equation 11 conditions the efficiency bias in the allocation of the budget to successful outcomes of the monitoring random variable $\theta_{i,t}$. Thus, the government tends to be more inquisitive with policymakers whose inefficiencies have been spotted in the past.

A.3 Indicator dynamics

We model indicator dynamics through a random growth process. Let γ_i denote a probability associated with the growth process experienced by indicator i. This probability depends on a combination of network effects (i.e., incoming spillovers) and budgetary allocations. Therefore, the growth process is modeled as independent Bernoulli trials with a probability of success

$$\gamma_{i,t} = \beta \frac{C_{i,t} + \frac{1}{n}\sum_j C_{j,t}}{1 + e^{-S_{i,t}}}, \tag{12}$$

where β is a normalizing parameter and $S_{i,t}$ are the net amount of spillovers received by indicator i in time t (this could be positive or negative). The spillovers are computed every period according to $S_{i,t} = \sum_j \mathbf{1}_{j,t}\mathbb{A}_{j,i}$, where $\mathbf{1}$ is the indicator function: 1 if indicator j grew in the previous period and 0 otherwise.

Next, we define the difference equation of indicator i as

$$I_{i,t+1} = I_{i,t} + \alpha_i \xi(\gamma_{i,t}) \tag{13}$$

where $\xi(\cdot)$ is the binary outcome (0 or 1) of a growth trial.

B Data

B.1 Sustainable development report

The first dataset is the one on development indicators provided by the 2020 Sustainable Development Report (SDR), which can be downloaded here: `github.com/sdsna/SDR2020`. These data contain 79 indicators for 166 countries. The sampling period is 2000 to 2020. While the SDR team has done substantial work minimizing the amount of missing observations, there are still some data gaps. Yet, its coverage is substantially better than the one of the UN SDG database. Table B.1 shows the number of observations per group (geographical cluster), and the percentages of missing observations that need to be interpolated or extrapolated. We explain the correction column ahead.

Table B.1: Summary of data preparation

Group	Observations	Interpolations (%)	Extrapolations (%)	Corrections (%)
Africa	44,142	7.16	26.63	16.94
E. Europe & C. Asia	22,575	5.55	24.76	17.30
East & South Asia	21,420	6.36	25.23	15.62
LAC	23,961	5.92	25.15	16.03

MENA	15,771	6.07	26.63	18.71
West	43,890	5.49	27.08	15.87

The SDR divides countries into groups. We re-define some of these groups to provide more intuitive results and to exploit group-level information in the data-imputation procedure. More specifically, we move Mexico and Chile from the OECD group to *LAC*. Likewise, Turkey is transferred from OECD to *East Europe & Central Asia*. Japan and Korea are reassigned from OECD to *East & South Asia*. The remaining countries in the OECD from Europe and North America are reassigned to a new group called *West*. Any remaining countries from Oceania are reassigned to *East & South Asia*. Countries with less than one million inhabitants are removed from the sample due to their large amount of missing observations. Cuba and North Korea are also dropped due to difficulties to exploit data from 'similar' countries in the data-imputation procedure. Finally, additional countries that do not have observations in the complementary datasets are also removed from the sample. This gives us a final sample size of 140 countries. Table B.2 provides the complete list of the countries, and their distribution across the six clusters. Table B.3 contains the codebook of the indicators.

Table B.2: List of countries by group

Group	Countries
Africa	AGO BDI BEN BFA BWA CAF CIV CMR COG ERI GAB GHA GIN GMB GNB KEN LBR LSO MDG MLI MOZ MRT MUS MWI NAM NER NGA RWA SDN SEN SLE SWZ TCD TGO TZA UGA ZAF ZMB ZWE
E. Europe & C. Asia	AFG ALB ARM AZE BGR BIH BLR CYP GEO HRV KAZ KGZ MDA MKD RUS TJK TKM TUR UKR UZB
East & South Asia	BGD CHN IDN IND JPN KHM KOR LAO LKA MMR MNG MYS NPL PAK PHL SGP THA VNM
LAC	ARG BOL BRA CHL COL CRI DOM ECU GTM HND HTI JAM MEX NIC PAN PER PRY SLV URY VEN
MENA	ARE BHR DZA EGY IRN IRQ JOR KWT LBN MAR OMN QAT SAU TUN
West	AUS AUT BEL CAN CHE CZE DEU DNK ESP EST FIN FRA GBR GRC HUN IRL ISR ITA LTU LVA NLD NOR NZL POL PRT SVK SVN SWE USA

Table B.3: List of countries by group

SDG	Code	Description
1	320pov	Poverty headcount ratio at \$3.20/day (%)
1	oecdpov	Poverty rate after taxes and transfers (%)
2	crlyld	Cereal yield (tonnes per hectare of harvested land)
2	obesity	Prevalence of obesity, BMI $\geq$ 30 (% of adult population)
2	snmi	Sustainable Nitrogen Management Index (worst 0-1.41 best)
2	trophic	Human Trophic Level (best 2-3 worst)
2	undernsh	Prevalence of undernourishment (%)
2	wasting	Prevalence of wasting in children under 5 years of age (%)
3	births	Births attended by skilled health personnel (%)
3	fertility	Adolescent fertility rate (births per 1,000 adolescent females aged 15 to 19)
3	hiv	New HIV infections (per 1,000 uninfected population)
3	incomeg	Gap in self-reported health status by income (percentage points)
3	lifee	Life expectancy at birth (years)
3	matmort	Maternal mortality rate (per 100,000 live births)
3	ncds	Age-standardized death rate due to cardiovascular disease, cancer, diabetes, or chronic respiratory disease in adults aged 30–70 years (%)
3	smoke	Daily smokers (% of population aged 15 and over)
3	swb	Subjective well-being (average ladder score, worst 0-10 best)
3	tb	Incidence of tuberculosis (per 100,000 population)
3	traffic	Traffic deaths (per 100,000 population)
3	u5mort	Mortality rate, under-5 (per 1,000 live births)
3	uhc	Universal health coverage (UHC) index of service coverage (worst 0-100 best)
3	vac	Percentage of surviving infants who received 2 WHO-recommended vaccines (%)
4	earlyedu	Participation rate in pre-primary organized learning (% of children aged 4 to 6)
4	pisa	PISA score (worst 0-600 best)
4	primary	Net primary enrollment rate (%)
4	second	Lower secondary completion rate (%)
4	socioec	Variation in science performance explained by socio-economic status (%)
4	tertiary	Tertiary educational attainment (% of population aged 25 to 34)
5	edat	Ratio of female-to-male mean years of education received (%)
5	familypl	Demand for family planning satisfied by modern methods (% of females aged 15 to 49 who are married or in unions)
5	lfpr	Ratio of female-to-male labor force participation rate (%)
5	parl	Seats held by women in national parliament (%)
5	paygap	Gender wage gap (% of male median wage)
6	safesan	Population using safely managed sanitation services (%)
6	safewat	Population using safely managed water services (%)
6	sanita	Population using at least basic sanitation services (%)
6	scarcew	Scarce water consumption embodied in imports (m^3/capita)
6	water	Population using at least basic drinking water services (%)
7	cleanfuel	Population with access to clean fuels and technology for cooking (%)
7	co2twh	CO_2 emissions from fuel combustion for electricity and heating per total electricity output ($MtCO_2$/TWh)
7	elecac	Population with access to electricity (%)
7	ren	Share of renewable energy in total primary energy supply (%)
8	accounts	Adults with an account at a bank or other financial institution or with a mobile-money-service provider (% of population aged 15 or over)
8	empop	Employment-to-population ratio (%)
8	impacc	Fatal work-related accidents embodied in imports (per 100,000 population)
8	unemp	Unemployment rate (% of total labor force)
8	yneet	Youth not in employment, education or training (NEET) (% of population aged 15 to 29)
9	articles	Scientific and technical journal articles (per 1,000 population)
9	intuse	Population using the internet (%)
9	lpi	Logistics Performance Index: Quality of trade and transport-related infrastructure (worst 1-5 best)
9	mobuse	Mobile broadband subscriptions (per 100 population)
9	netacc	Gap in internet access by income (percentage points)
9	patents	Triadic patent families filed (per million population)
9	rdex	Expenditure on research and development (% of GDP)
9	rdres	Researchers (per 1,000 employed population)
10	elder	Elderly poverty rate (% of population aged 66 or over)
10	palma	Palma ratio
11	pipedwat	Access to improved water source, piped (% of urban population)
11	pm25	Annual mean concentration of particulate matter of less than 2.5 microns in diameter (PM2.5) ($\mu g/m^3$)
11	rentover	Population with rent overburden (%)
11	transport	Satisfaction with public transport (%)
13	co2import	CO_2 emissions embodied in imports (tCO_2/capita)
13	co2pc	Energy-related CO_2 emissions (tCO_2/capita)
14	cleanwat	Ocean Health Index: Clean Waters score (worst 0-100 best)
14	cpma	Mean area that is protected in marine sites important to biodiversity (%)
14	fishstocks	Fish caught from overexploited or collapsed stocks (% of total catch)
14	trawl	Fish caught by trawling (%)
15	cpfa	Mean area that is protected in freshwater sites important to biodiversity (%)
15	cpta	Mean area that is protected in terrestrial sites important to biodiversity (%)
15	redlist	Red List Index of species survival (worst 0-1 best)
16	cpi	Corruption Perception Index (worst 0-100 best)
16	detain	Unsentenced detainees (% of prison population)
16	homicides	Homicides (per 100,000 population)
16	prison	Persons held in prison (per 100,000 population)
16	rsf	Press Freedom Index (best 0-100 worst)
16	safe	Percentage of population who feel safe walking alone at night in the city or area where they live (%)
17	govex	Government spending on health and education (% of GDP)
17	govrev	Other countries: Government revenue excluding grants (% of GDP)
17	oda	For high-income and all OECD DAC countries: International concessional public finance, including official development assistance (% of GNI)

B.2 Complementary datasets

The model of Policy Priority Inference (PPI) has a component that accounts for two aspects of public governance: the quality of monitoring efforts and the quality of the rule of law (φ and τ respectively, in the model's semantics). As explained in [8], both components are introduced as parameters that affect the agents' incentives to increase or decrease their levels of inefficiency. The intuition behind these parameters is to provide a comparative metric of the different qualities of public governance across countries. Therefore, rather than being actual estimates of qualities, they are indicators reflecting relative qualities. We use the Worldwide Governance Indicators database, which can be obtained here: `info.worldbank.org/governance/wgi`. In particular, we obtain the indicators of *control of corruption*, reflecting the quality of the monitoring efforts by the central authority, and the one of *rule of law*, capturing the quality of institutions designed to reassure a law-abiding society. These data are normalized between 0 and 1 for all the values available across countries and years. Then, for the countries in the SDR sample, we compute the inter-temporal values of these two indicators for the period 2000-2020.

Table B.4: Statistics by cluster

Group	Countries	Indicators	SDGs	Gov Exp	Monitoring (%)	Rule of Law (%)
Africa	39	58	15	4,516	28.44	41.18
E. Europe & C. Asia	20	74	16	26,956	29.02	43.00
East & South Asia	18	77	16	155,571	36.31	52.06
LAC	20	76	16	39,497	35.51	45.48
MENA	14	59	15	26,055	40.07	52.89
West	29	74	16	230,248	73.03	83.70

Government expenditure is in millions of US constant dollars (base 2010).

We obtain government expenditure data in General Government Final Consumption Expenditure (USD constant base 2010), built by the World Bank and obtainable here: `data.worldbank.org/indicator/NE.CON.GOVT.KD`. The sources of these data are the World Bank National Accounts Data, and the OECD National Accounts Data Files. Several countries have missing observations in this dataset. Therefore, we compute the yearly average of each country with the available data and multiply it by 21 years (from 2000 to 2020) in order to represent a proxy for the spending capacity of each country during the sampling period. Table B.2 shows a summary of these data.

We employ trade and geographical data for the data-imputation procedure. This information comes from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII). Trade data on imports and exports between every country is provided by the CEPII TradeProd Database (covering 1980 to 2006), available here: `www.cepii.fr/CEPII/fr/bdd_modele/presentation.asp?id=5`. The information on geographical proximity weighted by urban population centers is obtained from the CEPII GeoDist Database (with coverage for 2005) and can be downloaded from: `www.cepii.fr/CEPII/en/bdd_modele/bdd_modele.asp`.

C Data pre-processing

C.1 Normalization

All the variables are normalized between 0 and 1. This is a standard procedure that not only allows an easier interpretation, but also helps on the performance of machine learning methods to impute missing observations. For a given indicator, all the data on all years and countries are pooled. Then, the formula $(I - I_{\min})/(I_{\max} - I_{\min})$ is applied to the pooled vector I, where $I_{\min}$ and $I_{\max}$ correspond to the theoretical minimum and maximums that the indicator values can take. $I_{\min}$ and $I_{\max}$ are provided by the SDR dataset. Once all the indicators have been normalized, we invert those where a reduction indicates improvement (e.g., poverty rate) by subtracting the normalized indicator from 1. This is also a standard procedure in the development literature. Once we have normalized the data, we proceed to impute the missing values.

C.2 Interpolation

A small percentage of the missing data can be imputed through interpolations. Standard procedures such as linear interpolations are common practice among analysts and consultants. We, instead, use the Gaussian Process Regression approach, which is acknowledged in the machine-learning community as a superior data imputation method because it finds the non-linear function that could predict the time series better. We employ the Python package `Scipy`, and its function `GaussianProcessRegressor` on each individual time series.

C.3 Extrapolation

A more challenging problem is imputing missing observations that lie before/beyond the first/last observed data point of a time series. For the interpolation task, we consider that the country-specific observations provide enough information to estimate the Gaussian process because, in general, the imputed values are in line with the observed levels of volatility, so the imputations are unlikely to exceed the theoretical minimum and maximum values. However, for missing observations that are not surrounded by data points, a Gaussian process estimated only on an individual time series could yield extrapolations that violate the theoretical minimums and maximums of the indicators (and constrained Gaussian processes is still a young research field). Thus, imputation via extrapolation could greatly benefit from the information contained in time series of the same indicator, but from other countries. Of course, the other countries to be taken as reference data should be selected based on certain similarity criteria. We construct as similarity index for this purpose.

For a given country i, the similarity index to another country j takes into account:

- If both countries share a common border ($border_{i,j}$);
- if they belong to the same country group ($group_{i,j}$);
- their distance, weighted by population centers ($distance_{i,j}$);
- the total imports of i from j ($imports_{ij}$) and;
- the total exports from i to j ($exports_{i,j}$).

The variable $border_{i,j}$ is binary and takes the value 1 if there is a shared border, and 0 otherwise. Component $group_{i,j}$ is also binary and becomes 1 if both countries belong to the same group (i.e., geographical cluster) and 0 otherwise. The term $distance_{i,j}$ is the geographical distance provided by the CEPII dataset between i and j, divided by the largest distance between i and any other country, and subtracted from 1. The value of $imports_{ij}$ consists of the total number of imports received by country i from i, divided by the maximum number of imports received by i from any country. Similarly, $exports_{ij}$ consists of the total number of exports sent by country i to j, divided by the maximum number of exports sent by i to any country. Finally, the similarity index is expressed as

$$similarity_{i,j} = border_{i,j} + group_{i,j} + distance_{i,j} + imports_{ij} + exports_{ij}.$$

We compute the similarity index for every pair of countries. Then, for a given country i, we rank all other countries according to the index. We select the top 3 most similar countries to i, and create a pooled dataset that includes these nations and i (i.e., a reference group composed by 4 countries in total). Since not all countries have the same indicators, it is possible that, for country i the reference group may be different from one indicator to another. Once the the pooled dataset of a single indicator has been built, we perform the extrapolation procedure for country i only. Notice that, since every country has a different geography, the reference groups across countries are likely to be unique, which helps preserving some contextual information about country i during the extrapolation.

In order to extrapolate the time series of a given indicator in a specific country, we use the the Multi-Output Gaussian Process Toolkit [10], which is publicly available here: `games-uchile.github.io/mogptk/`. This is a method that combines Gaussian processes with more sophisticated machine learning algorithms, such as artificial neural networks, to exploit the information about the same indicator in all members of the reference group, while balancing its relevance with respect to the data of country i. Because of the uniqueness of the reference groups across countries and indicators, this estimation has to be performed individually for each row in the database, meaning that we have to fit more than 9000 models. Despite the computational-intensive nature of this procedure, it is the preferred one as it yields more reliable extrapolations than any other method commonly used by development analysts (the method performs cross-validation algorithms to fit the Gaussian process).

C.4 Correction

In spite of the virtues of the Multi-Output Gaussian Process Toolkit to produce reliable extrapolations, a fraction of them may still lie beyond the theoretical limits of the indicators. To correct the estimations, we perform a variance compression procedure that preserves the periodicity of the extrapolations, but re-normalizes the data in order to bound them to the limits established in the

SDR dataset. We apply this procedure also to those extrapolations that, even if they remain within the theoretical bounds, their variance exceed the empirical one.[1] By correcting the variance of the extrapolations, we produce data imputations with a volatility that is closer to the empirical one.

To explain the variance compression procedure, let us consider forward extrapolations. Given a time series $I_{2000,t}$ covering the years $\{2000, \ldots, t\}$ and an extrapolation $E_{t+1,2020}$ covering $\{t + 1, \ldots, 2020\}$, we want to compress the extrapolation such that $var(E_{t+1,2020}) \leq var(I_{2000,t})$. We perform this compression in a procedural fashion by iteratively re-normalizing $E_{t+1,2020}$ by a factor $z \lessapprox 1$. The compression procedure for forward extrapolations is described in algorithm 2.

Algorithm 2: Variance compression pseudocode

Input: $I_{2000,t}$ and $E_{t+1,2020}$

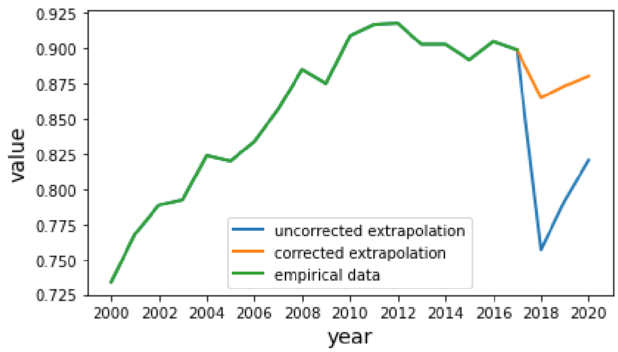
while $var(E_{t+1,2020}) > var(I_{2000,t})$ *or any value in* $E_{t+1,2020}$ *lies beyond a theoretical limit* **do**

  $E_{t+1,2020} = I_{2000,t}(t) + z[E_{t+1,2020} - I_{2000,t}(t)]$;

Figure C.1 shows an example of the outcome of this procedure. In this case, the extrapolation remains within the theoretical boundaries, but the variance was substantially larger with respect to the one from the empirical time series. This variance may have been the result of large variations in the time series of the reference group. Thus, with the compression algorithm we are able to preserve the information on relative fluctuations and trend direction provided by the reference group, while normalizing the imputed data to be consistent with the empirical one. The same logic and algorithm applies to backward extrapolations. The fraction of observations that were subjected to this procedure is reported in Table B.1.

[1] A variance beyond the one exhibited by a specific indicator in a given country is to be expected in any extrapolation procedure that pools data from other countries.

Figure C.1: Example of variance compression



C.5 Instrumental indicators

In PPI, policy priorities (budgetary allocations) are defined over those indicators that are considered to have a direct impact through specific expenditure programs; that is why we call them *instrumental indicators*. In this study, we identify a subset of indicators that, from our experience, are likely to be instrumental. The rest of the indicators are defined as *collateral* because we find them too aggregate for any government to claim any capability of direct manipulation. Of course, some indicators could be instrumental in some countries but not in others. This, however, requires extensive contextual knowledge, something difficult to obtain when studying 140 nations. Therefore, the 55 indicators identified in Table C.1 are assumed to be instrumental in all countries from our sample.

Table C.1: Instrumental indicators

SDG	Indicator code
1	320pov
1	oecdpov
2	snmi
2	undernsh
2	wasting
3	births
3	incomeg
3	matmort

3	tb
3	u5mort
3	uhc
3	vac
4	earlyedu
4	pisa
4	primary
4	second
4	socioec
4	tertiary
5	edat
5	familypl
6	safesan
6	safewat
6	sanita
6	water
7	cleanfuel
7	co2twh
7	elecac
7	ren
8	accounts
8	yneet
9	intuse
9	lpi
9	mobuse
9	netacc
9	rdex
10	elder
11	pipedwat
11	pm25
11	rentover
11	transport
13	co2import
13	co2pc
14	cleanwat
14	cpma

14	fishstocks
14	trawl
15	cpfa
15	cpta
15	redlist
16	detain
16	homicides
16	prison
16	safe
17	govex
17	govrev

D Network

D.1 Estimation

The network of interlinkages consists of a directed acyclic graph estimated through Bayesian methods from the package `sparsebn` [14], which can be accessed here: `https://github.com/itsrainingdata/sparsebn`. The links do not represent causal relationship, but conditional dependencies. This is discussed in detail in [8, 15]. The method works well with short time series, yet it benefits from using longer or repeated time series. Therefore, we employ the pooled data from the referenced groups assembled during the extrapolation procedure. The method assumes no temporal dependence, so the observations for the same indicator across different countries can be concatenated in a unified series. Before this concatenation, we transform the individual series into their first differences to remove the effect of temporal trends. In this manner, a unique dataset with 80 observations (of first differences) for each indicator is produced for each country, so the estimated networks are context-specific.

Another virtue of `sparsebn` is its ability to specify a 'white list' of edges that can be considered true positives. In other words, with prior knowledge, one can determine a set of links that would be expected from the estimation. We identify 109 synergies (links with positive weights) that should be expected in any network of any country. These synergies are reported in Table D.1.

Table D.1: White list of synergies

Origin	Destination
wpc (SDG 1)	undernsh (SDG 2)
wpc (SDG 1)	u5mort (SDG 3)
wpc (SDG 1)	fertility (SDG 3)
wpc (SDG 1)	vac (SDG 3)
wpc (SDG 1)	primary (SDG 4)
wpc (SDG 1)	earlyedu (SDG 4)
wpc (SDG 1)	accounts (SDG 8)
wpc (SDG 1)	netacc (SDG 9)
wpc (SDG 1)	elder (SDG 10)
320pov (SDG 1)	undernsh (SDG 2)
320pov (SDG 1)	u5mort (SDG 3)
320pov (SDG 1)	fertility (SDG 3)
320pov (SDG 1)	vac (SDG 3)
320pov (SDG 1)	primary (SDG 4)
320pov (SDG 1)	earlyedu (SDG 4)
320pov (SDG 1)	accounts (SDG 8)
320pov (SDG 1)	netacc (SDG 9)
320pov (SDG 1)	elder (SDG 10)
oecdpov (SDG 1)	undernsh (SDG 2)
oecdpov (SDG 1)	u5mort (SDG 3)
oecdpov (SDG 1)	fertility (SDG 3)
oecdpov (SDG 1)	vac (SDG 3)
oecdpov (SDG 1)	primary (SDG 4)
oecdpov (SDG 1)	earlyedu (SDG 4)
oecdpov (SDG 1)	accounts (SDG 8)
oecdpov (SDG 1)	netacc (SDG 9)
oecdpov (SDG 1)	elder (SDG 10)
undernsh (SDG 2)	u5mort (SDG 3)
undernsh (SDG 2)	lifee (SDG 3)
undernsh (SDG 2)	swb (SDG 3)
undernsh (SDG 2)	pisa (SDG 4)
wasting (SDG 2)	ncds (SDG 3)

wasting (SDG 2)	lifee (SDG 3)
obesity (SDG 2)	ncds (SDG 3)
obesity (SDG 2)	lifee (SDG 3)
trophic (SDG 2)	obesity (SDG 2)
crlyld (SDG 2)	undernsh (SDG 2)
snmi (SDG 2)	crlyld (SDG 2)
matmort (SDG 3)	oecdpov (SDG 1)
matmort (SDG 3)	lifee (SDG 3)
matmort (SDG 3)	swb (SDG 3)
neonat (SDG 3)	lifee (SDG 3)
u5mort (SDG 3)	lifee (SDG 3)
tb (SDG 3)	u5mort (SDG 3)
ncds (SDG 3)	swb (SDG 3)
fertility (SDG 3)	second (SDG 4)
births (SDG 3)	matmort (SDG 3)
births (SDG 3)	u5mort (SDG 3)
vac (SDG 3)	u5mort (SDG 3)
uhc (SDG 3)	oecdpov (SDG 1)
uhc (SDG 3)	u5mort (SDG 3)
uhc (SDG 3)	tb (SDG 3)
uhc (SDG 3)	ncds (SDG 3)
uhc (SDG 3)	vac (SDG 3)
uhc (SDG 3)	swb (SDG 3)
incomeg (SDG 3)	oecdpov (SDG 1)
smoke (SDG 3)	ncds (SDG 3)
smoke (SDG 3)	lifee (SDG 3)
primary (SDG 4)	swb (SDG 3)
second (SDG 4)	edat (SDG 5)
second (SDG 4)	yneet (SDG 8)
pisa (SDG 4)	empop (SDG 8)
socioec (SDG 4)	pisa (SDG 4)
science (SDG 4)	pisa (SDG 4)
resil (SDG 4)	pisa (SDG 4)
familypl (SDG 5)	fertility (SDG 3)
edat (SDG 5)	fertility (SDG 3)
edat (SDG 5)	lfpr (SDG 5)

edat (SDG 5)	paygap (SDG 5)
lfpr (SDG 5)	parl (SDG 5)
lfpr (SDG 5)	paygap (SDG 5)
water (SDG 6)	u5mort (SDG 3)
water (SDG 6)	swb (SDG 3)
sanita (SDG 6)	u5mort (SDG 3)
sanita (SDG 6)	swb (SDG 3)
elecac (SDG 7)	empop (SDG 8)
cleanfuel (SDG 7)	co2pc (SDG 13)
co2twh (SDG 7)	co2pc (SDG 13)
ren (SDG 7)	cleanfuel (SDG 7)
unemp (SDG 8)	intuse (SDG 9)
empop (SDG 8)	intuse (SDG 9)
empop (SDG 8)	mobuse (SDG 9)
empop (SDG 8)	govrev (SDG 17)
yneet (SDG 8)	empop (SDG 8)
intuse (SDG 9)	accounts (SDG 8)
lpi (SDG 9)	empop (SDG 8)
rdex (SDG 9)	rdres (SDG 9)
rdres (SDG 9)	articles (SDG 9)
rdres (SDG 9)	patents (SDG 9)
netacc (SDG 9)	empop (SDG 8)
adjgini (SDG 10)	rentover (SDG 11)
palma (SDG 10)	rentover (SDG 11)
pipedwat (SDG 11)	water (SDG 6)
transport (SDG 11)	swb (SDG 3)
rentover (SDG 11)	swb (SDG 3)
co2pc (SDG 13)	ncds (SDG 3)
co2import (SDG 13)	ncds (SDG 3)
cpma (SDG 14)	fishstocks (SDG 14)
cpma (SDG 14)	trawl (SDG 14)
cpta (SDG 15)	redlist (SDG 15)
cpfa (SDG 15)	fishstocks (SDG 14)
safe (SDG 16)	swb (SDG 3)
cpi (SDG 16)	homicides (SDG 16)
cpi (SDG 16)	safe (SDG 16)

prison (SDG 16)	detain (SDG 16)
govex (SDG 17)	uhc (SDG 3)
govex (SDG 17)	tertiary (SDG 4)
govex (SDG 17)	rdex (SDG 9)
govrev (SDG 17)	govex (SDG 17)

We identify 4 trade-offs (links with negative links) that should be expected. Trade-offs are substantially less than synergies because they convey highly contextual information. We report them in table D.1.

Table D.2: White list of trade-offs

Origin	Destination
elecac (SDG 7)	co2twh (SDG 7)
elecac (SDG 7)	co2pc (SDG 13)
empop (SDG 8)	pm25 (SDG 11)
empop (SDG 8)	co2pc (SDG 13)

The specification of the white list does not force the estimation to yield a specific sign. Instead, `sparsebn` takes the white lists and forces the algorithm to maintain those links in the estimated network. It may be the case that some of these links come out with the opposite sign from the expected one. We consider these to be false positives so we remove these links from the network in a refinement procedure.

D.2 Refinement

Besides eliminating links with an incorrect sign, we also remove negative edges between indicators that belong to the same SDG. The intuition is that trade-offs are likely to occur across topics in different SDGs, not in the same one. While there is still the possibility of intra-SDG trade-offs, we rather sacrifice them and allow this type of error (losing some true positives) than permitting a large amount of false positives.

Finally, it is still possible that certain links have excessively large magnitudes in their weights, i.e. outliers. We consider these to be false positives as such magnitudes are likely to be an artifact of the data. To eliminate these links, we establish weight thresholds in the 5 and 95 percentiles of the weights of all the networks pooled together. If the weight of a particular link lies below or above these thresholds, it is eliminated from its corresponding network.

E Calibration nuances

The are some detail regarding the calibration procedure that should be clarified for reproducible research. First, while the indicators are normalized between 0 and 1, if their values are too close to zero, Brent's optimization method may have difficulties finding optimal values. To solve this, it is necessary to multiply the indicators by a scalar. We find that re-scaling by 100 solves any issue.

Second, for simplicity, let us express the goals vector $T_i, \ldots, T_N$ as the difference between the goals minus the initial values. In the case of a retrospective estimation, that would be the final empirical values minus the initial ones. Thus, the initial values for every indicator are always zero, and the goals are the gaps to be closed. This transformation has no effect in the results since, in contrast to [9], the levels of the indicators do not affect the model, only their changes.

Third, since the model produces non-negative growth dynamics in the indicators, the goals vector needs to have always positive values, but it may be the case that some indicators produce negative ones because their final values are below the initial ones. In [8], this was solved by a data-shifting procedure. Here, refine this principle to consider the best-performing period of an indicator. If indicator i has a goal $T_i < 0$, we replace it by $\max(I_{i,\cdot}) - I_{i,0}$. This correction implies that a lower final value may have been the result of an exogenous event, and that the maximum value that the indicator had in the sampling period is a better proxy to the actual achievements of the country in this policy issue. If the goal is still negative (e.g., because the indicator persistently worsened), then we establish the goal $T_i = 10^{-3}$ after having re-scaled the indicators. In these way, the worst-performing indicators will exhibit the lowest historical progress and, thus, the lowest estimated structural factors. Further nuances could be introduce to distinguish levels of worst-performing indicators. However, the interpretation of such details seem rather arbitrary and context specific, something we rather avoid for this study.

E.1 Calibrated parameters

This appendix presents the average parameters structural factors estimated for each country and their corresponding errors. Due to the numerous parameters calibrated, it is unfeasible to report each one on a table, so we provide their country-level means. However, the complete datasets with all the parameters and their errors can be obtained from the repository `github.com/oguerrer/SDG_feasibility`.

Table E.1: Calibrated parameters and estimation errors

Country	Average α	Average error from α	Parameter β	Average error from β
AGO	0.3743	0.0017	0.0000	0.0010
BDI	0.4334	0.0018	0.0000	0.0012
BEN	0.3139	0.0021	0.0000	0.0029
BFA	0.4126	0.0015	0.0000	0.0015
BWA	0.4343	0.0019	0.0000	0.0040
CAF	0.2885	0.0014	0.0000	0.0004
CIV	0.4505	0.0024	0.0000	0.0019
CMR	0.3837	0.0017	0.0000	0.0024
COG	0.5138	0.0024	0.0000	0.0025
ERI	0.3037	0.0012	0.0000	0.0004
GAB	0.3997	0.0021	0.0000	0.0480
GHA	0.4359	0.0020	0.0000	0.0036
GIN	0.3272	0.0015	0.0000	0.0063
GMB	0.3583	0.0016	0.0000	0.0465
GNB	0.3325	0.0012	0.0000	0.0003
KEN	0.4642	0.0025	0.0000	0.0010
LBR	0.3050	0.0010	0.0000	0.0019
LSO	0.6474	0.0075	0.0000	0.0031
MDG	0.2760	0.0013	0.0000	0.0054
MLI	0.3831	0.0024	0.0000	0.0045
MOZ	0.4002	0.0024	0.0000	0.0007
MRT	0.3335	0.0017	0.0000	0.0049
MUS	0.2635	0.0014	0.0000	0.0064
MWI	0.3390	0.0021	0.0000	0.0019
NAM	0.7862	0.0045	0.0000	0.0025

NER	0.3344	0.0018	0.0000	0.0035
NGA	0.3069	0.0017	0.0000	0.0012
RWA	0.5359	0.0024	0.0000	0.0022
SDN	0.4690	0.0018	0.0000	0.0045
SEN	0.4457	0.0014	0.0000	0.0005
SLE	0.3244	0.0019	0.0000	0.0020
SWZ	0.3351	0.0011	0.0000	0.0007
TCD	0.3218	0.0021	0.0000	0.0003
TGO	0.4077	0.0023	0.0000	0.0007
TZA	0.4485	0.0023	0.0000	0.0033
UGA	0.3444	0.0015	0.0000	0.0028
ZAF	0.3170	0.0015	0.0000	0.0025
ZMB	0.5075	0.0024	0.0000	0.0015
ZWE	0.3835	0.0024	0.0000	0.0003
AFG	0.4609	0.0023	0.0000	0.0006
ALB	0.5325	0.0014	0.0000	0.0042
ARM	0.3866	0.0020	0.0000	0.0007
AZE	0.4736	0.0019	0.0000	0.0025
BGR	0.4418	0.0022	0.0000	0.0052
BIH	0.3472	0.0015	0.0000	0.0004
BLR	0.7266	0.0046	0.0000	0.0025
CYP	0.3294	0.0014	0.0000	0.0012
GEO	0.3928	0.0019	0.0000	0.0007
HRV	0.3485	0.0016	0.0000	0.0018
KAZ	0.3859	0.0014	0.0000	0.0006
KGZ	0.3287	0.0011	0.0000	0.0021
MDA	0.3236	0.0017	0.0000	0.0010
MKD	0.4648	0.0043	0.0000	0.0037
RUS	0.2974	0.0020	0.0000	0.0004
TJK	0.3300	0.0015	0.0000	0.0037
TKM	0.2550	0.0012	0.0000	0.0063
TUR	0.2819	0.0018	0.0000	0.0035
UKR	0.3265	0.0015	0.0000	0.0002
UZB	0.3401	0.0015	0.0000	0.0033
BGD	0.4270	0.0017	0.0000	0.0063
CHN	0.3637	0.0020	0.0000	0.0042

IDN	0.3529	0.0024	0.0000	0.0003
IND	0.4580	0.0014	0.0000	0.0010
JPN	0.1953	0.0008	0.0000	0.0043
KHM	0.4471	0.0018	0.0000	0.0008
KOR	0.2446	0.0014	0.0000	0.0054
LAO	0.4427	0.0017	0.0000	0.0055
LKA	0.3588	0.0019	0.0000	0.0176
MMR	0.3566	0.0024	0.0000	0.0030
MNG	0.4224	0.0022	0.0000	0.0012
MYS	0.2537	0.0009	0.0000	0.0025
NPL	0.5368	0.0022	0.0000	0.0031
PAK	0.2968	0.0010	0.0000	0.0003
PHL	0.2679	0.0016	0.0000	0.0044
SGP	0.2590	0.0006	0.0000	0.0061
THA	0.3301	0.0017	0.0000	0.0015
VNM	0.3914	0.0025	0.0000	0.0013
ARG	0.2726	0.0013	0.0000	0.0029
BOL	0.3829	0.0015	0.0000	0.0002
BRA	0.2721	0.0017	0.0000	0.0033
CHL	0.3252	0.0016	0.0000	0.0002
COL	0.3280	0.0021	0.0000	0.0011
CRI	0.3539	0.0016	0.0000	0.0016
DOM	0.4215	0.0025	0.0000	0.0012
ECU	0.3579	0.0012	0.0000	0.0028
GTM	0.3290	0.0015	0.0000	0.0014
HND	0.3338	0.0011	0.0000	0.0017
HTI	0.2639	0.0012	0.0000	0.0037
JAM	0.2976	0.0021	0.0000	0.0016
MEX	0.2949	0.0013	0.0000	0.0018
NIC	0.3879	0.0014	0.0000	0.0004
PAN	0.3172	0.0017	0.0000	0.0052
PER	0.3395	0.0016	0.0000	0.0025
PRY	0.3516	0.0014	0.0000	0.0013
SLV	0.4660	0.0020	0.0000	0.0005
URY	0.2682	0.0010	0.0000	0.0060
VEN	0.3719	0.0017	0.0000	0.0052

ARE	0.5013	0.0023	0.0000	0.0005
BHR	0.3690	0.0012	0.0000	0.0020
DZA	0.3522	0.0024	0.0000	0.0067
EGY	0.2987	0.0017	0.0000	0.0013
IRN	0.4653	0.0015	0.0000	0.0003
IRQ	0.3835	0.0021	0.0000	0.0056
JOR	0.3257	0.0015	0.0000	0.0008
KWT	0.3171	0.0017	0.0000	0.0001
LBN	0.2845	0.0011	0.0000	0.0022
MAR	0.4123	0.0015	0.0000	0.0498
OMN	0.3765	0.0016	0.0000	0.0015
QAT	0.3774	0.0018	0.0000	0.0015
SAU	0.5094	0.0020	0.0000	0.0020
TUN	0.3576	0.0012	0.0000	0.0009
AUS	0.2398	0.0012	0.0000	0.0010
AUT	0.2535	0.0011	0.0000	0.0043
BEL	0.2724	0.0013	0.0000	0.0049
CAN	0.1969	0.0011	0.0000	0.0019
CHE	0.2323	0.0010	0.0000	0.0051
CZE	0.3268	0.0016	0.0000	0.0006
DEU	0.2322	0.0013	0.0000	0.0018
DNK	0.2039	0.0007	0.0000	0.0010
ESP	0.3323	0.0010	0.0000	0.0035
EST	0.4044	0.0014	0.0000	0.0020
FIN	0.2064	0.0008	0.0000	0.0025
FRA	0.2591	0.0014	0.0000	0.0011
GBR	0.2216	0.0008	0.0000	0.0011
GRC	0.3372	0.0022	0.0000	0.0026
HUN	0.3587	0.0011	0.0000	0.0022
IRL	0.2949	0.0009	0.0000	0.0012
ISR	0.2422	0.0021	0.0000	0.0016
ITA	0.2661	0.0013	0.0000	0.0018
LTU	0.4016	0.0015	0.0000	0.0004
LVA	0.3661	0.0018	0.0000	0.0036
NLD	0.2321	0.0009	0.0000	0.0006
NOR	0.2499	0.0007	0.0000	0.0156

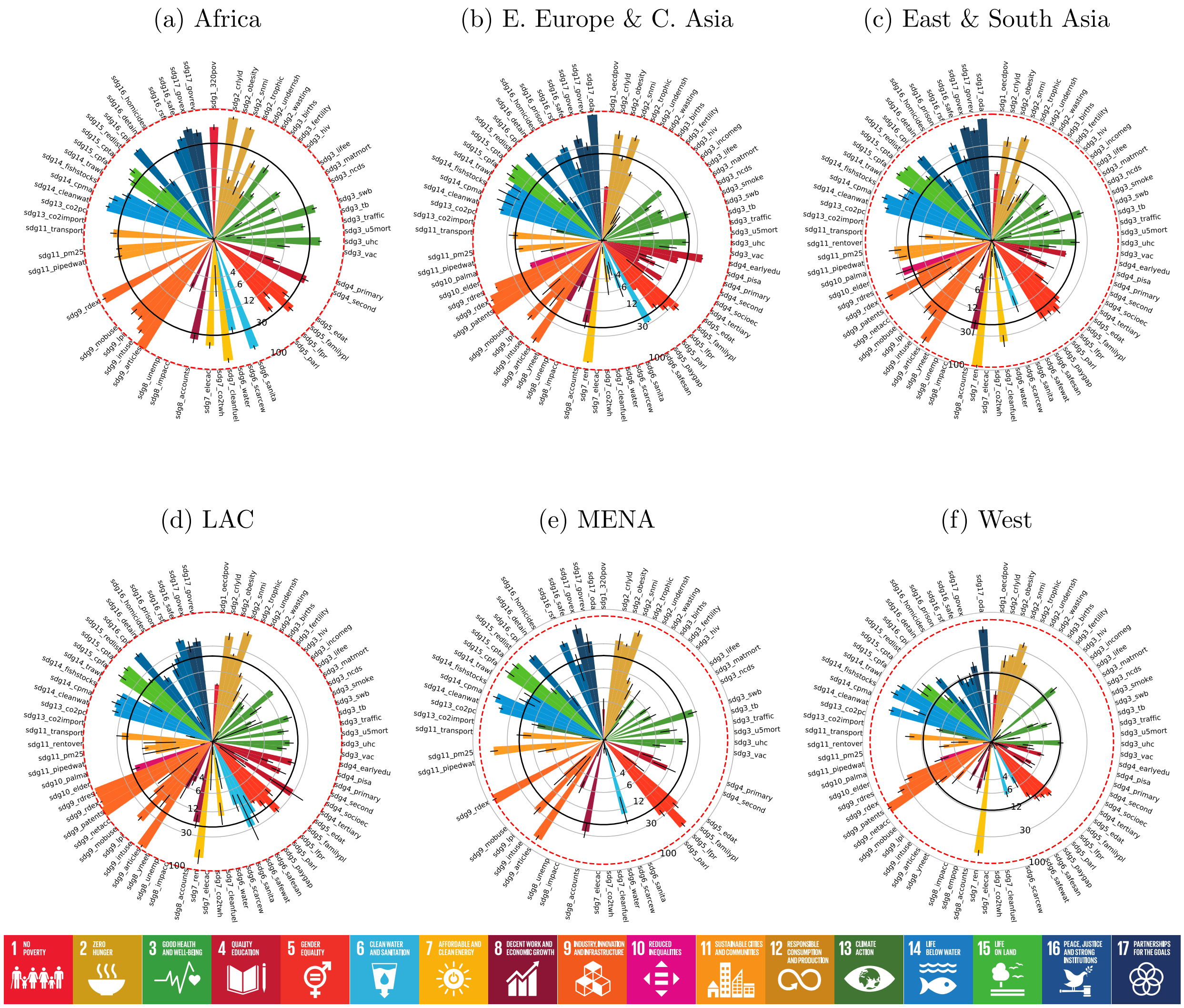
NZL	0.3008	0.0018	0.0000	0.0017
POL	0.3125	0.0019	0.0000	0.0043
PRT	0.3469	0.0013	0.0000	0.0016
SVK	0.3402	0.0016	0.0000	0.0010
SVN	0.3736	0.0015	0.0000	0.0019
SWE	0.2031	0.0011	0.0000	0.0027
USA	0.1979	0.0007	0.0000	0.0004

Due to the numerous parameters α calibrated, this table reports the average value for α in each country, and the average error associated to the optimized α. It also reports parameter β and its associated error. The value of β is always reported as zero because of the monetary units of the data, which implies setting a very small value for β in order to generate probability values in [0,1] for γ.

F Robustness to the disbursement schedule

In the calibration procedure we assume that all the indicators reach their final values in $\mathcal{T} = 50$ simulation periods (the disbursement schedule). This implies that the disbursement schedule is being mapped to the number of years covered in the dataset. However, there exists the possibility that the simulation results could be biased by this assumption. Hence, as a robustness test, we modify the number of disbursement periods to 25 and analyze if there are significant changes in the gaps estimated for 2030. When comparing Figure 2 in the main text and Figure F.1 below, we notice that this is not the case.

Figure F.1: Average gap by 2030 using 25 (instead of 50) disbursement tics



Height of each bar: average gap between the SDG and the indicator level predicted by 2030 computed across countries in the cluster. Empty spaces between bars: no data was available for the corresponding indicator in any country from the cluster. Solid black ring: average gap across across countries (in the cluster) and indicators. Dashed red ring: largest average gap (between indicators in the cluster). Black lines at the top of each bar: ± standard error of the mean gaps across the countries of a cluster.

G Results for shorter time series

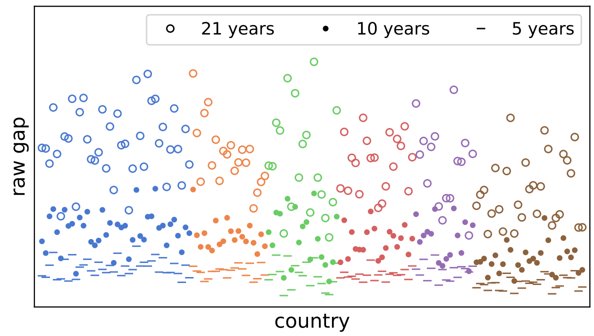
Another interesting robustness test has to do with analyzing the consequences of reducing the number of years in the sampling period (i.e., reducing the observations to a more recent period). We explore this by using the most recent 5 or 10 years, instead of the 21 included in the database. With these reduced datasets, we re-estimate the network of interlinkages and re-calibrate the model. For a more recent sample, one would expect a shorter historical gap, and smaller calibrated structural factors $\alpha_1, \ldots, \alpha_N$. Structural factors are not directly comparable across simulations that use a different disbursement schedule or sampling periods of different lengths. In order to take into account these differences, comparisons need to be done between the SDG gaps estimated for 2030. Panel a in figure G.1 shows the average historical gap for each country for the different levels of coverage in the datasets. Panel b shows the same gaps but divided by the number of years of coverage in the respective datasets. As expected, once normalized by the length of the time series, the gaps show no systematic differences. This indicates that there were no substantial/systematic improvements in the rhythm of development during the last decade or the last 5 years. Panel c confirms that the estimated structural factors are smaller for the reduced dataset. Finally, panel d shows the average indicator difference between the three datasets for each country. That is, the average difference (in absolute value) between the estimated indicators levels for 2030 of two different forms of the dataset (21, 10, and 5 years). Clearly, there are differences but they are modest, not exceeding 5% in any country. When translating these results into gaps and aggregating them into regions and SDGs, the differences become negligible, yielding very similar results to the ones presented in the main text.

If there were substantial structural improvements in later years, these should be discernible through significantly smaller SDG gaps for the reduced datasets. If that is not the case, it means that the our assumption of capturing long-term structural factors in $\alpha_1, \ldots, \alpha_N$ is reasonable since the data does not show abrupt accelerations in short time spans.

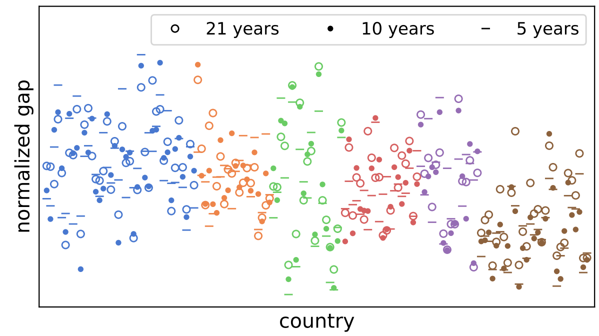
The result of our estimations with the reduced datasets are described in Figures G.2 and G.3 for 10 and 5 years, respectively. They indicate that, indeed, the structural factors did not change in the short- and medium-run, hence they capture long-term dynamics. In fact, the results are strikingly similar for the average gaps across countries within each group.

Figure G.1: Historical gaps and structural factors

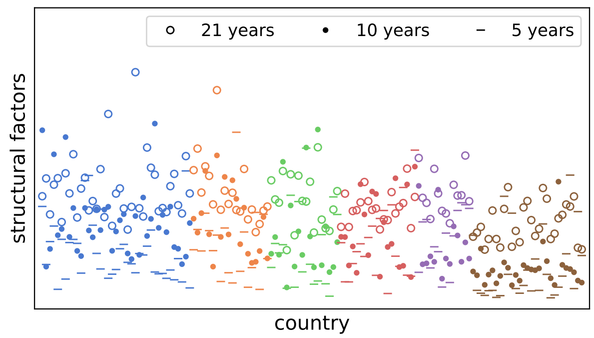
(a) Raw gaps



(b) Normalized gaps



(c) Structural factors



(d) Average indicator difference

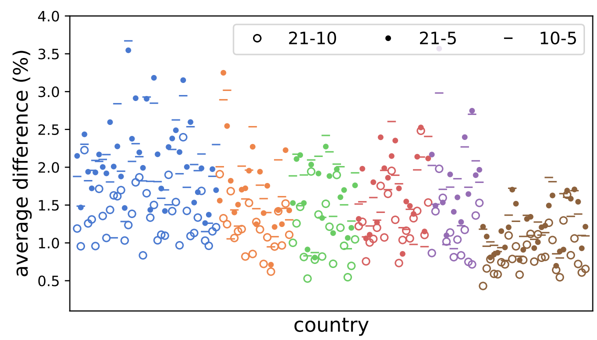
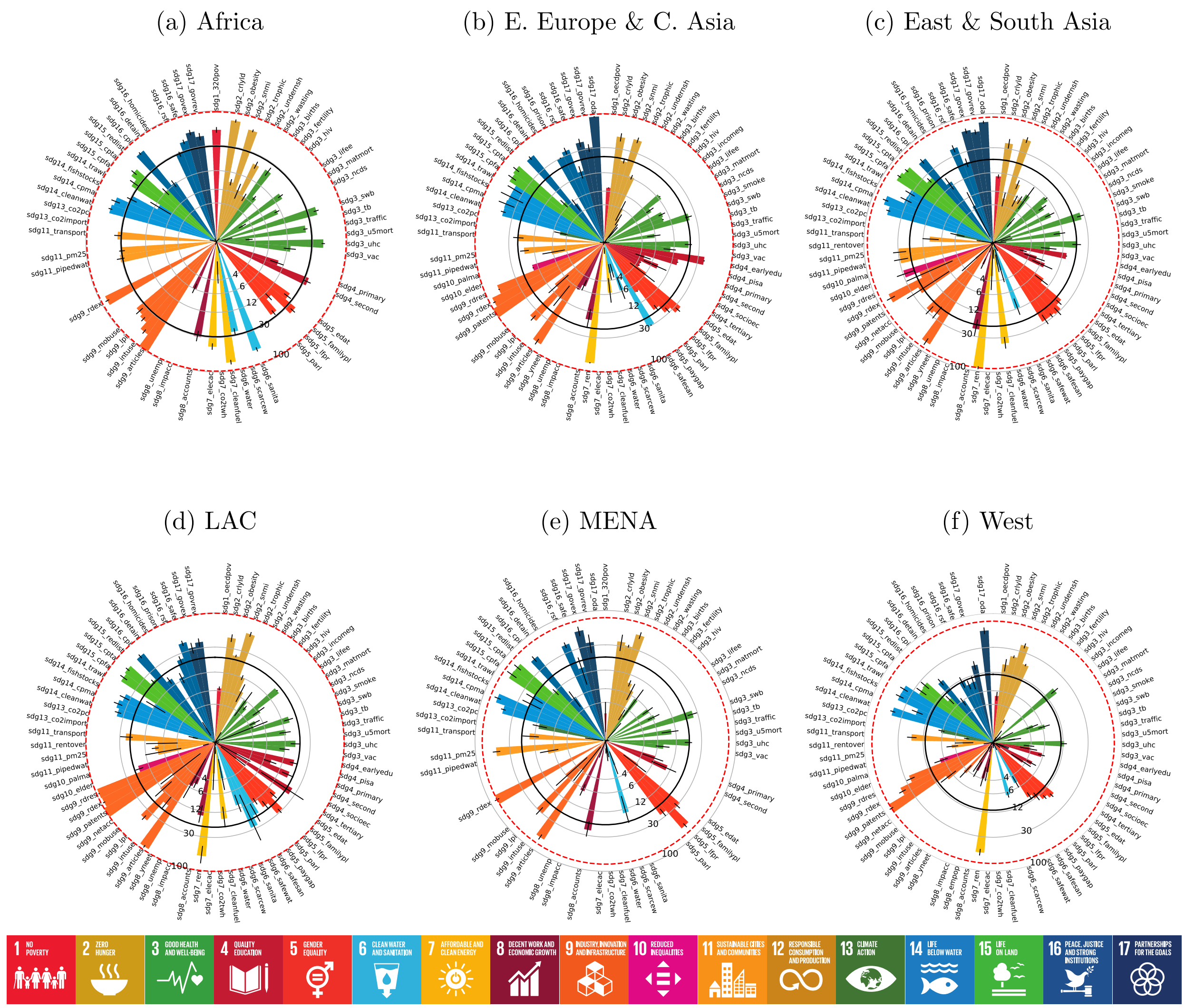


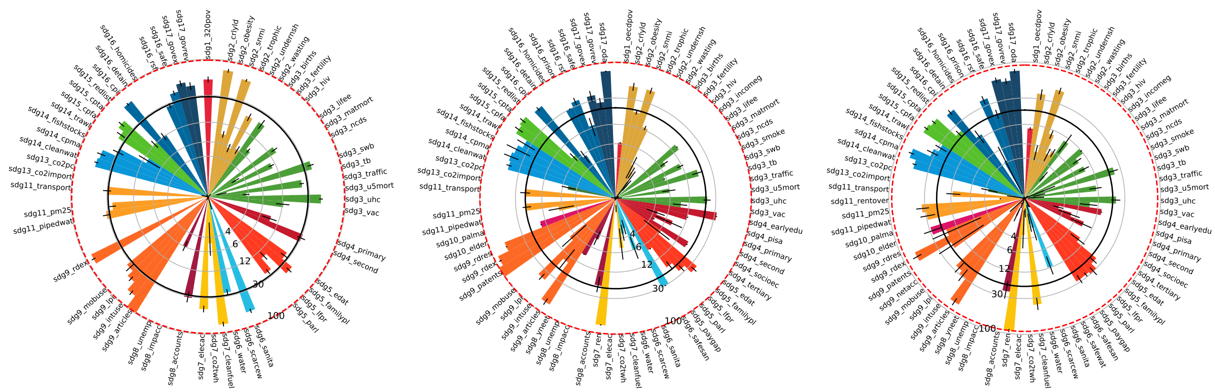
Figure G.2: Average gap by 2030 using 10 years of data



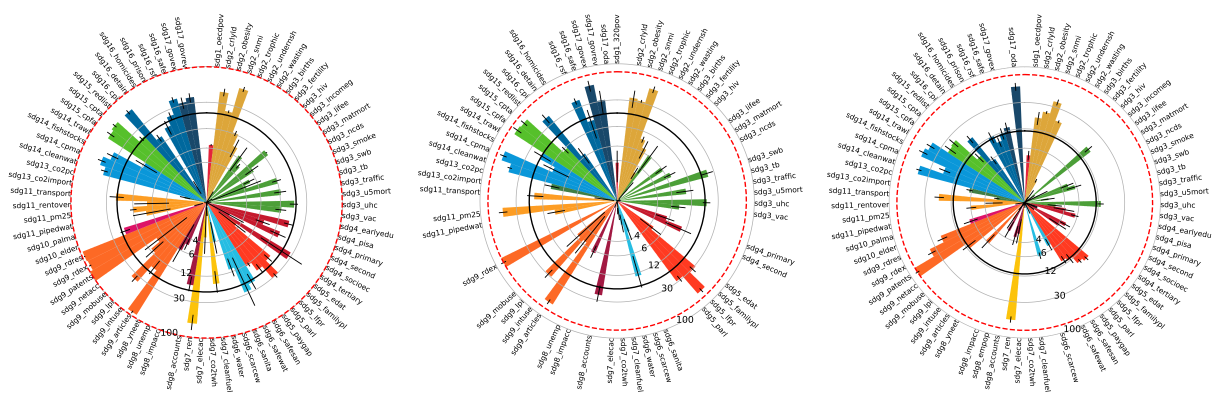
Height of each bar: average gap between the SDG and the indicator level predicted by 2030 computed across countries in the cluster. Empty spaces between bars: no data was available for the corresponding indicator in any country from the cluster. Solid black ring: average gap across across countries (in the cluster) and indicators. Dashed red ring: largest average gap (between indicators in the cluster). Black lines at the top of each bar: ± standard error of the mean gaps across the countries of a cluster.

Figure G.3: Average gap by 2030 using 5 years of data

(a) Africa (b) E. Europe & C. Asia (c) East & South Asia



(d) LAC (e) MENA (f) West



Height of each bar: average gap between the SDG and the indicator level predicted by 2030 computed across countries in the cluster. Empty spaces between bars: no data was available for the corresponding indicator in any country from the cluster. Solid black ring: average gap across across countries (in the cluster) and indicators. Dashed red ring: largest average gap (between indicators in the cluster). Black lines at the top of each bar: ± standard error of the mean gaps across the countries of a cluster.