Policy Priority Inference for Sustainable Development

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

[1]The Alan Turing Institute

[2]University College London

[3]Tecnológico de Monterrey Campus Puebla

Abstract

We develop a computational framework to support the planning and evaluation of development strategies towards the 2030 Agenda. The methodology takes into account the complexities of the political economy underpinning the policymaking process, for example, the multidimensionality of development, the interlinkages between these dimensions, the inefficiencies of implementing policy interventions, as well as the institutional factors that promote or disencourage these inefficiencies. The framework is scalable and usable with publicly-available development-indicator data, and it can be further refined as more data becomes available, for example, on public expenditure. We demonstrate its usage through an application for the Mexican federal government. For this, we infer historical policy priorities, *i.e.* non-observable allocations of transformative resources that generate changes in development indicators. We also show how to use the tool to assess the feasibility of development goals, to measure policy coherence, and to identify accelerators. Overall, the tool provides a well-rounded framework that allows policymakers and other stakeholders to embrace a complexity view to tackle the challenges of the Sustainable Development Goals.

Acknowledgements

We are grateful for the sponsorship of the United Nations Development Programme, Bureau for Latin America (UNDPLAC) and the UNDP-Mexico. We would like to thank Luis Felipe López Calva and Almudena Fernández from UNDPLAC for supporting this initiative; to Florian Chávez Juárez, Georgina Trujillo, Lucy Hackett and Alejandro Blasco from the National Laboratory of Public Policy for their support with data preparation and logistics; and Annabelle Sulmont, Stephanus Visser and Maite García from the UNDP-Mexico for their involvement in the organisation of the project and the workshops. We are also grateful with all the participants from the various seminars, conferences and workshops where PPI has been presented; in particular, the seminar series on PPI conducted at the UNDP in New York and the two workshops at the National Laboratory of Public Policy in Mexico City.

Disclaimer

While this research has been supported by the UNDP, the opinions and interpretations expressed in this paper reflect the authors' viewpoint only, not the the ones of UNDP nor from any other individual and organization involved in developing the project.

1 Introduction

In its 2030 Agenda Declaration (UN General Assembly, 2015, p.2), the United Nations acknowledge the importance of understanding development as a process with numerous dimensions interacting with each other: "*The interlinkages and integrated nature of the SDGs are of crucial importance in ensuring that the purpose of the new Agenda is realized.*". While the multidimensional aspect of development has been extensively discussed in the development literature (McGillivray and Shorrocks, 2005; Chambers and Institute of Development Studies (Brighton, 2007; Alkire and Foster, 2011), the complexity of development expressed through interlinkages is certainly a novel aspect of the latest international agendas. Acknowledging such complexity opens the door to new ways of thinking about development and, consequently, calls for suitable methodologies emanating from Complexity Science and from Computational Social Science. Given that these principles are ingrained at the core of the Sustainable Development Goals (SDGs), the design of development strategies for the 21$^{\text{st}}$ needs to be open to paradigm shifts if it truly wants to embrace complexity and achieve its ambitious objectives.

This paper introduces introduces a new framework to study policy prioritization through the lens of Complexity Science, using computational tools. This framework was developed through a project sponsored by the United Development Programme (UNDP), conducted between March and December of 2019 in Mexico, and involving various stakeholders at the federal and state levels. As we show throughout the paper, the proposed framework allows tackling challenges that are central to the current discussion of the 2030 Agenda; for example, assessing the *feasibility* of development goals and identifying development *accelerators*.

The rest of the paper is structured in the following way. Section 2 discusses why a complexity approach is necessary to meet the 2030 Agenda. Section 3 provides a detailed explanation of the model underpinning PPI. In section 4, we present the data assembled for the UNDP project. Section 5 shows the results of a historical analysis. Then, in section 6, we present the outcome a various prospective analyses. Finally, section 7 concludes and discusses future directions.

2 On the complexity of development

While the highly-discussed networks of interlinkages reveal some of the complexity behind development, the veil remains at large because socioeconomic systems exhibit other types of intricacies; some of them already part of academic discussions and public discourse. Here, we mention a few that we find relevant from our own research and numerous interactions with policymakers and academics in the field. First, there is the problem of the inefficiencies that dampen the effectiveness of the resources spent in transformative policies. A typical symptom of such wastage is corruption; a problem that is part of public discourse and international agendas (Izquierdo et al., 2018; Baum et al., 2019).

Second, and related to corruption, we have the institutional dimension of development; in particular, public governance (Bank, 2017; IMF, 2019). Given that monitoring efforts and the strength of the rule of law shape the incentives of those who can profit from corruption, the institutions that facilitate public governance are key to the impact of development policies. Third, while corruption and public governance have traditionally been studied by economists from the principal-agent point of view (Rose-Ackerman, 1975; Klitgaard, 1988), today we know that the lack of a committed principal (a common feature in developing countries) leads to a collective-action problem (Persson et al., 2013), rendering traditional incentive-driven interventions ineffective. The complexity arising from collective-action dynamics –such as social norms– explains why reforms to the rule of law around the developing world have failed in curving corruption (Bank, 2017; Guerrero and Castañeda, 2019).

Fourth, in the macro scale of international agendas, causation between policy interventions and development outcomes is almost impossible to infer from available data sources such as development indicators.[1] This is so because policy interventions take

[1] While useful in highly-specific contexts at the micro-level, experimental procedures are not viable at the scale and level of aggregation at which these international agendas operate.

place at a micro-level, while development indicators are typically macro-level variables. Thus, in absence of highly-granular multi-level data, vertical causal mechanisms cannot be properly addressed by traditional statistical tools such as regression analysis (Casini and Manzo, 2016; Ospina-Forero et al., 2019). Instead, a production account of causation is a more adequate framework and, once again, Complexity Science has developed the relevant tools.

Fifth, since data on the amount of resources spent in specific development indicators is practically non-existent, a –commonly used– second best approach is assuming that some development indicators are exogenous variables. That is, analysts presuppose that the chosen "independent" variables can be directly intervened or manipulated. Consequently, they justify interventions in those policy issues if the associated regression coefficients are statistically significant. In reality, development indicators are the result of government actions that are partly motivated by the performance of the indicators themselves. Thus, the apparent "exogenous" change in a covariate might actually originate from policy outcomes associated to the dependent variable. Consider the example of a government whose policies successfully reduce the rate of tuberculosis among individuals under 15 years old. Suppose that these improvements are accompanied by a significant growth in literacy rates across he same population. Indeed, a healthier population can lead to better educational outcomes, so it is reasonable to justify a regression model where improvements in the literacy rate are explained by a reduction in tuberculosis (the chosen "exogenous" variable). Now, suppose that the government has been training teachers to improve the quality of education, and that this policy succeeds. Once the teachers have been trained, the government is able to shift resources towards strengthening tuberculosis diagnostics and treatments, producing the observed outcome. This is not the typical reverse-causality problem since there is no direct link from quality-of-education to tuberculosis. Instead, this endogeneity issue arises from a political economy process that is unobservable through development indicators. This poses severe limitations to data-driven analyses of aggregate development indicators. Thus, it is necessary to develop data-generating models where political-economy considerations –such as the reallocation of resources– are explicit. Computational Social Science provide us with tools to build such models.

The previous argument points to a political-economy aspect of development that lives in the very core of every strategy and international agenda: *policy prioritization*. Understanding how governments prioritize the allocation of resources across numerous policy issues is a prevalent and rarely accounted-for challenge that requires a complexity perspective. This paper introduces a computational tool that facilitates the analysis of policy priorities and other related development issues. The proposed framework models how governments allocate resources across different topics, with the objective of achieving a set of goals, in an environment characterized by the complexity features previously discussed. The method generates (bottom-up) development-indicator dynamics from these allocations. Thus, by matching these synthetic data to empirical indicators, this approach allows inferring the policy priorities that, arguably, were partly responsible for the observed indicator dynamics. Hence, we call it *Policy Priority Inference* (PPI).

PPI builds on a model developed by Castañeda et al. (2018), consisting of a central authority allocating resources across a multidimensional and interlinked policy space. The CCG model has been previously applied to the study of policy resilience (Castañeda and Guerrero, 2018), ex-ante policy evaluation (Castañeda and Guerrero, 2019), policy coherence (Guerrero and Castañeda, ming), and corruption (Guerrero and Castañeda, 2019). However, its ability to be used to provide detailed policy advice is limited by problems such as not being able to account for negative interlinkages between indicators (which excludes environmental issues) and the need to build international datasets for calibration purposes. Despite these limitations, the CCG model offers a first approach to study policy prioritization from a complexity perspective. Therefore, improving it to make it suitable for the SDGs seem like a natural next step.

The application of PPI presented in this paper is an early proof of concept, not an evaluation nor an assessment of how well the government is doing. In order to produce evaluations of this sort, it is necessary to improve the input data, something that we are currently working on in collaboration with different adopters of the tool, but that falls out of the scope of this paper. Throughout the paper, we elaborate on novel ways in which PPI can be used to address salient concepts of the 2030 Agenda such as *policy coherence for sustainable development*, *accelerators*, and *SDG budgeting*. By demonstrating the usefulness of PPI in tackling these problems, we hope to make a strong case for why the PPI philosophy should become an integral part of any government's planning toolkit.

3 Model

Like any other model, the one developed here contains certain assumptions, simplifications and limitations; therefore, it is natural to think that more detailed and context-relevant models could be constructed in the future. For this reason, PPI should be thought of as a philosophy rather than as a specific model. The distinctive feature of such philosophy is that, in order to understand policy prioritization, a production account of causation is necessary (Casini and Manzo, 2016). In other words, it is not enough to study dependencies between aggregate variables, but to model the micro-level processes that give raise to aggregate dynamics. This is so because, ultimately, policy priorities lead to micro-level policy interventions. These interventions, in turn, generate aggregate outcomes in non-linear and non-trivial ways. Therefore, the PPI philosophy advocates for analytic tools that stem from Complexity Science and Computational Social Science. Here, we adopt one such tool: agent-computing, popularly known as agent-based modeling. In the rest of this section, we explain the model in great detail, and in a bottom-up fashion. Firstly, we elaborate on the behavior of policymaking agents. Second, we present the government's strategy. Third, we connect these two elements to the aggregate dynamics of the development indicators. Due to space limitations, we provide further technical details on model calibration and robustness tests in the online Appendix D.[2]

[2] In addition, the public repository `github.com/oguerrer/PPI` provides open source code, data and tutorials on how to use PPI and to replicate the results presented in this paper.

3.1 Micro-foundations 1: inefficiencies

3.1.1 Public servant benefits

Let us assume that there are n agents, each in charge of a public policy that is specific to a single policy issue.[3] In order to implement the mandated policy in a given period t, agent i receives $P_{i,t}$ resources from the central authority. With these resources, the policymaker tries to leverage two potential benefits: (1) the reputation from being a proficient public servant and (2) the utility derived from being inefficient. Proficiency, on the one hand, is beneficial because it signals competence to the central authority and to the political system. Therefore, proficient agents gain political status that may catapult their careers in the future. Inefficiency, on the other hand, is also beneficial because it appeals to the utility derived from private gains. That is, by devoting time and resources to other activities such as shirking, diverting funds, or benefiting friends (through corrupt public bids), an agent may substitute the benefits from proficiency with the private gains of becoming inefficient. Of course, there is no free lunch in becoming inefficient, as the central authority may exert monitoring activities and take punitive measures with the purpose of increasing proficiency. The effectiveness of such mechanisms, however, is bound to the institutional setting of each nation, and we elaborate on that later.

We formalize the trade-off between proficiency and inefficiency through

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

where $F_{i,t+1}$ represents the benefit or utility obtained in the next period.

The first summand of equation 1 captures the benefit from 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

[3] An agent may well represent a specific bureaucrat, a minister, or a government agency.

$$\Delta I^*_{i,t} = \frac{I_{i,t} - I_{i,t-1}}{\sum_j |I_{j,t} - I_{j,t-1}|}, \tag{2}$$

an it captures the idea that the central authority evaluates the performance of each policy through development indicators.

Going back to the first summand of equation 1, 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. As we will show ahead, $0 \leq C_{i,t} \leq P_{i,t}$, so the factor $\frac{C_{i,t}}{P_{i,t}}$ represents the efficiency with which resources are being used in policy issue i.

Next, let us focus on the second addend of equation 1, 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. We previously mentioned that 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 the literature of public governance, θ and τ represent two fundamental institutional factors: the *quality of monitoring efforts* and the *strength of the rule of law*.

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

$$\lambda_{i,t} = \varphi \frac{1}{1 + e^{-D_{i,t}}}. \tag{3}$$

Parameter φ in equation 3 corresponds to the quality of the monitoring efforts. Note that both $\varphi \in [0, 1]$ and $\tau \in [0, 1]$, and that they are time- and indicator-independent. This means that these parameters can be directly calibrated from empirical data such as development indicators of public governance.[4]

The second factor that determines whether inefficiencies are spotted in equation 3 is $D_{i,t}$. This represents the level of the private gain extracted by agent i, relative to the private gains of all other agents. Formally, this quantity is obtained from

$$D_{i,t} = \frac{(P_{i,t} - C_{i,t}) - \min(P_{\cdot,t} - C_{\cdot,t})}{\max(P_{\cdot,t} - C_{\cdot,t}) - \min(P_{\cdot,t} - C_{\cdot,t})} - \frac{1}{2}, \tag{4}$$

where the term -1/2 is necessary to specify a balanced logistic function in equation 3.

Our motivation to correlate the probability of being spotted with the relative level of inefficiency is rather intuitive. Large inefficiencies such as corruption scandals come into the spotlight when they stand out from the norm. Thus, in contrast with the traditional principal-agent view, PPI considers the collective-action problem of social norms that prevent the principal from aligning the agents' incentives.

Now that we have established how the benefit function in equation 1 works, it is clear that the task of the agent is to determine the level of contribution $C_{i,t}$. Since agents face an environment with uncertainty and, as we will show ahead, there are interdependencies between the indicators, we adopt a robust and empirically validated reinforcement learning model: *directed learning* (Dhami, 2016).

3.1.2 Public servant learning

The principle behind directed learning is that actions can go in one of two directions: positive or negative; and outcomes reward or disencourage future actions in the same direction. For example, if an agent became more inefficient and, then, his/her benefits increased, then s/he will become even more inefficient the next period. If, in contrast, the government was able to penalize the agent so that his/her benefits decreased, s/he would become more proficient the next period. Formally, action $X_{i,t}$ of agent i can be modeled as

[4] An alternative approach where both parameters are endogenous and time-dependent can be found in Castañeda et al. (2018).

$$X_{i,t+1} = X_{i,t} + \text{sgn}((X_{i,t} - X_{i,t-1})(F_{i,t} - F_{i,t-1}))|F_{i,t} - F_{i,t-1}|, \tag{5}$$

where sgn($\cdot$) is the sign function.

In reality, $X_{i,t}$ is an abstraction of any type of action that an agent may take in order to be inefficient (*e.g.*, shirking, diverting funds, favoring friends, etc.), so X may have any real value. In order to map action $X_{i,t}$ into the amount of effective resources, we define

$$C_{i,t} = \frac{P_{i,t}}{1 + e^{-X_{i,t}}}. \tag{6}$$

Equation 6 incorporates the directed learning model into the policymaking process, while making sure that $C_{i,t} \leq P_{i,t}$. This completes the micro-foundations that give place to a specific type of inefficiencies: *technical inefficiencies*, *i.e.* those arising from the policymaking process. This part of the model has no free parameters (since φ and τ can be obtained from data). Therefore, the learning model does not require any calibration procedure.

3.2 Micro-foundations 2: government policy priorities

The policy priorities are represented by the allocation profile $P = P_i, \ldots P_n$. At this point, 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 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 of financial development. Alternatively, an indicator may be collateral simply because it is not relevant to the government priorities, so there are no dedicated policies. Naturally, policy priorities can only be defined on the n instrumental indicators, while there can only be n public servants (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.

While a government determines its policy priorities over the n instrumental nodes, it may have aspirations to improve the N indicators, even without explicit policy instruments for the collateral ones (political campaigns that promise improvements in issues that are out of the politician's control are a good example). Such aspirations are captured in a vector of *targets* or *goals* (we use there two terms interchangeably) $T_0, \ldots, T_N$. Note that we have established a clear difference between goals and priorities, two concepts that are often confused in the literature of sustainable development. Goals, on the one hand, represent the aspirations of a government. They are the exogenous variables in PPI and consist of specific values what the central authority wants to reach for each indicator. Priorities, on the other hand, are not aspirations, but actions. Because of the political economy of the policymaking process, priorities are endogenous variables. Hence, the endogeneity problem arising from policy prioritization is directly address in PPI.

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

$$\min \left[\sum_{i}^{N} (T_i - I_{i,t})^2 \right] \tag{7}$$

in every period through different allocations. The central authority achieves this through its allocation profile P. Note that this formulation implies that the government wants to achieve goals for topics in which it may not necessarily have policy instruments. This is why concepts such as *accelerators*[5] are so important to the 2030 Agenda.

What determines the distribution of resources P? Different countries and their governments may have various motivations for allocating different resources across policy issues. For example, a welfare state may be more welcoming of pro-social policies such as unemployment benefits and social housing, while a technological oriented one may prioritize R&D investment. PPI is flexible enough to allow any function or algorithm

[5] Informally, accelerators are policy issues whose improvement catalyze the development of other topics via spillovers.

to model how a government steers its priorities. This, of course, requires certain priors and data. In absence of such information, we remain agnostic about the specificity of each government and provide a simple yet non-trivial policy prioritization heuristic. First, the government agent uses the rule of prioritizing laggard topics because there is a generalized belief that these issues are bottlenecks to development. In fact, this was the approach in the Millennium Development Goals (gar, 2017, p.1). Hence, the government measures the normalized gaps between targets and indicators

$$G_{i,t} = \frac{(T_i - I_{i,t}) - \min(T_{\cdot} - I_{\cdot,t})}{\max(T_{\cdot} - I_{\cdot,t}) - \min(T_{\cdot} - I_{\cdot,t})}. \tag{8}$$

Besides supporting poorly developed issues, we assume that governments avoid systematically allocating resources to ineffective policies. Therefore, our government heuristic also takes into account the normalized history of spotted inefficiencies of each agent, represented by

$$H_{i,t} = \frac{\sum_l^t \theta_{i,l}(P_{i,l} - C_{i,l}) - \min\left[\sum_l^t \theta_{\cdot,l}(P_{\cdot,l} - C_{\cdot,l})\right]}{\max\left[\sum_l^t \theta_{\cdot,l}(P_{\cdot,l} - C_{\cdot,l})\right] - \min\left[\sum_l^t \theta_{\cdot,l}(P_{\cdot,l} - C_{\cdot,l})\right]}. \tag{9}$$

The target-indicator gaps encourage prioritization, while a reputation of inefficiency discourages it. However, the international experience suggests that the gaps play a more central role than the inefficiencies (Izquierdo et al., 2018). Thus, we look for a function where the allocation $P_{i,t}$ depends mainly on $G_{i,t}$ and, in a lesser measure, on $H_{i,t}$. We adopt the functional form.

$$q_{i,t} = {G_{i,t}}^{1+H_{i,t}}, \tag{10}$$

which can be normalized to obtain the policy priorities

$$P_{i,t} = \frac{q_{i,t}}{\sum_j q_{j,t}}. \tag{11}$$

Equation 10 is rather intuitive. In term $G_{i,t}^{1+H_{i,t}}$, the base is always fractional, while the exponent is always greater than one. This means that policy issues with more visible inefficiencies will be penalized more. Appendix D.5 shows that the estimated allocation profiles are robust across different functional forms of equation 10.

3.3 Macro-dynamics: development indicators

Now that we have built the micro-foundations of the model, we connect them to the evolution of the development indicators. In doing so, we aim at generating indicator dynamics that resemble empirical data in three aspects: (1) that each indicator starts at a given value and reaches specific final value, (2) that all indicators arrive to their final values at the same time, and (3) that these data presents a certain level of volatility. We address attribute (1) here, and leave (2) and (3) for the calibration procedure, which are explain in Appendix D.

3.3.1 Indicator dynamics

From the micro foundations, we know that a fraction $C_{i,t}$ (the contribution) of an allocation $P_{i,t}$ is efficiently used in public policies. In conjunction with the incoming spillover effects $S_{i,t}$ (these could be positive and negative), public policies transform the associated indicator $I_{i,t}$. We model this transformation through a random growth process. Let α_i denote the amount of growth experienced in indicator i, conditional on a successful outcome. By successful outcome we mean that the event of growing at a rate α_i is a random variable where the success rate depends on incoming spillovers and public policies (if any). Therefore, the growth process is modeled as independent Bernoulli trials with probability of success

$$\gamma_{i,t} = \frac{\alpha_i + C_{i,t}/P_t^*}{\alpha_i + e^{-\frac{NS_{i,t}}{\sum_j (T_j - I_{j,t})/(T_j - I_{j,0})}}}, \tag{12}$$

where P_t^* is the maximum amount of allocated resources across all policy issues in period t. Note that $\frac{C_{i,t}}{P_t^*} = \frac{C_{i,t}}{P_{i,t}} \frac{P_{i,t}}{P_t^*}$, which means that the effect of the contribution is the combination of how efficiently are the resources being used ($C_{i,t}/P_{i,t}$) and how much resources policy issue i receives in relation to all other issues ($P_{i,t}/P_t^*$). The sum dividing the spillover term is a correction that we explain ahead together with α_i.

Next, we define the growth equation of indicator i as

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

where $\xi(\cdot)$ is the binary outcome (0 or 1) of a growth trial. The growth factor α_i lives in $(0, 1)$, and it shapes both the probability of success and the amount of growth. We can think of α_i as all the other determinants responsible for an indicator's growth, but that are not explicit in the model.[6] This factor needs to be calibrated in order to match the second attribute of the data: all indicators reach their final values at the same time. Note that the gap $T_i - I_{i,t}$ shrinks as the indicator grows, and $I_{i,t} < T_i$ in every period because α_i. This means that the indicator is guaranteed to converge to T_i, which takes care of the first attribute of the data: reaching the indicator's specific final value. Because of these logistic-like convergence dynamics, the change in the indicator becomes, on average, smaller as it approaches T_i, so the magnitude of spillover effects decreases with time. In order to correct this artifact, equation 12 divides the spillover term in the exponent by the average normalized gap at time t.

Note that the probabilistic nature of equation 13 is consistent with the fact that indicators are not directly manipulable, but rather, governments try to affect them through policy interventions, which some times succeed and some time fail, depending on the efficiency and efficacy of their implementation. It also implies that the interlinkages do not represent causal relations, but conditional dependencies that increase or decrease the chance of improving an indicator. This is consistent with our earlier point on the impossibility to establish causation between aggregate indicators. We elaborate on the impossibility of causal networks of development indicators in the next section.

[6]An example of the factors that α capture is infrastructure. A good road network is an important factor to improve domestic trade. If the infrastructure is good enough, trade can grow at a "natural" rate without direct interventions (at least as long as the road capacity can sustain it). Here, the α corresponding to domestic trade contains the influence of this type of infrastructure. Of course, α may contain other factors that facilitate domestic trade, for example, income per capita and technological readiness. Hence, an experiment that considers structural reforms would be to modify α to understand its effect on reaching a set of goals.

3.3.2 SDG networks and spillovers

Let us define a network with N nodes, each one corresponding to an indicator. An arrow $i \rightarrow j$ represents a change on indicator j conditioned on a change by indicator i, not a causal link. That is, the existence of $i \rightarrow j$ means that, if we observe a change in j, a change in i was likely to have taken place. However, a change in i do not necessarily trigger a change in j (otherwise it would be a causal link).[7] In terms of the model, a positive edge $i \rightarrow j$ indicates a higher likelihood of j growing, while a negative one translates in a lower likelihood. This is consistent with conditional dependencies. A further discussion of why SDG networks –by themselves– cannot capture causal relations is provided by Ospina-Forero et al. (2019).

We say that a spillover from i to j takes place through the interaction of i's change $\Delta I_{i,t} = I_{i,t} - I_{i,t-1}$ and the intensity of the conditional dependency specified in the adjacency matrix $\mathbb{A}$. Therefore, the incoming spillovers from i to j in period t are

$$S_{i \rightarrow j,t} = \Delta I_{i,t-1} \mathbb{A}_{ij}, \tag{14}$$

which can be positive, negative or zero. We are interested in the amount of incoming spillovers that each node receives. Thus, the relevant measure to consider is the net incoming spillovers

$$S_{j,t} = \sum_i \Delta I_{i,t-1} \mathbb{A}_{ij}, \tag{15}$$

which is one of the determinants of successful growth in equation 12.

3.4 Summary

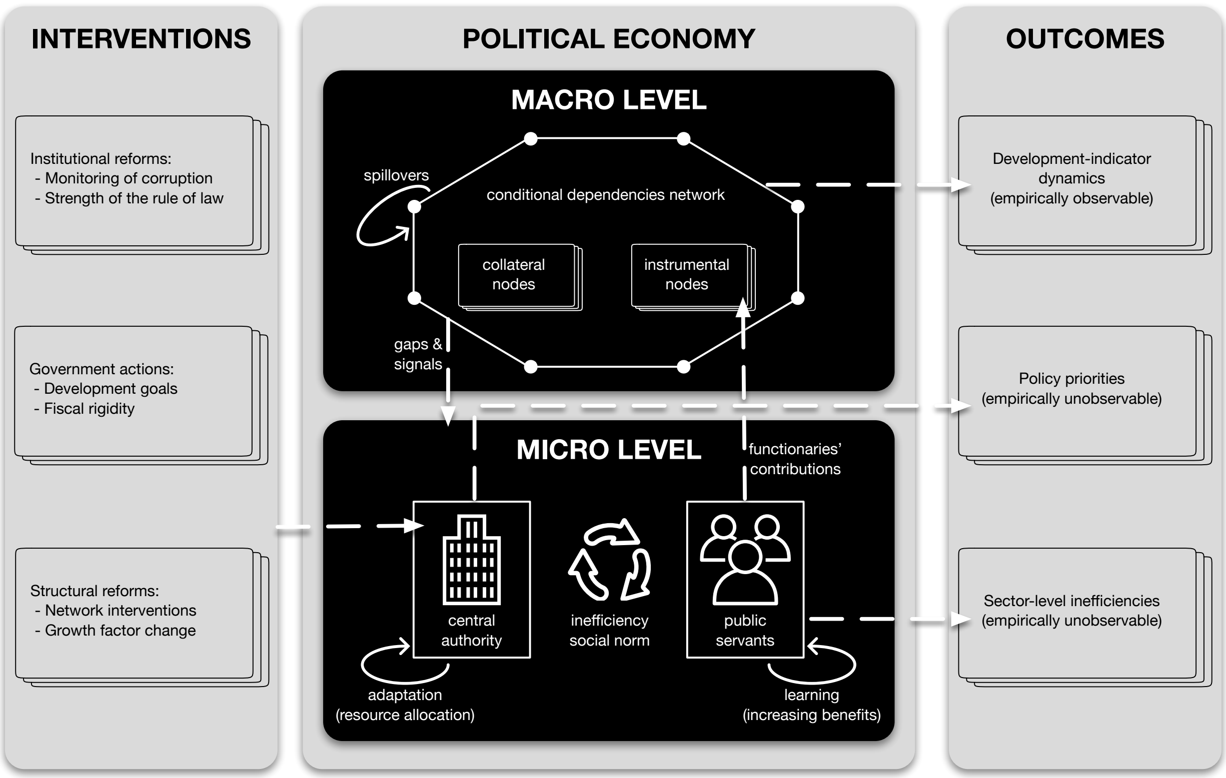
Figure 1 provides a summary diagram of the model. The left panel presents some topics that can be considered exogenous interventions in PPI, for example, reforms to the judicial power to strengthen the rule of law (modifying the relevant parameter), aligning

[7] Conditional dependencies are not plain correlations either. A correlation is just a co-movement of two variables, but which could be produced by a third variable, so no conditioning between i and j is necessary.

the countries development goals to the SDGs (changing the target vector T), or promoting synergistic relations between sectors (re-weighting interlinkages). Note that all the interventions take place at the micro-level, so they are relevant to the government agent. Of course, in the long run, every aspect of development could be considered endogenous. Here, we have made an assumption regarding what we consider exogenous. In our view, for example, things like development goals are a much more exogenous than indicators. This is so because the goals represent the government's aspirations, and these, in turn, come from much broader processes such as societal consensus and international agreements. Endogenizing the formation of government aspirations implies modeling political processes and working in a long-term horizon. Having said this, we should point out that, in contrast to most existing analytic tools that support the 2030 Agenda, PPI focuses on short-term analysis. That is, PPI is not designed to forecast indicator levels in 30 or 50 years, but to aid in assessing the feasibility of reaching development goals within a few of government terms. Thus, PPI is most useful to national and sub-national governments designing development plans; treasuries planning budgets; political parties crafting political campaigns; consultants and grading agencies assessing a government's commitment to development; NGO's evaluating development strategies; multilateral organizations coordinating international agendas; etc.

Appendix B provides a list with all the model variables. As a summary, we would like to point out that PPI takes four exogenous sources of information as inputs: (1) initial conditions, (2) spillover network, (3) targets or goals, and (4) governance parameters. The first one is usually collected by governments and international organizations. The second one can be estimated via quantitative or qualitative methods (see Ospina-Forero et al. (2019) for a review on network-estimation methods). The third is an exogenous variable that can be built from societal consensus, political platforms, public consultations, etc. The fourth can be obtained from international datasets such as the Worldwide Governance Indicators. Algorithm 1 summarizes the model in a few lines.

Figure 1: Structure of the PPI model



Algorithm 1: PPI pseudocode

Input: $\alpha_1, \ldots, \alpha_N$, initial I, T, $\mathbb{A}$, φ, τ

while $T_i - I_i > \varepsilon$ *for every* i **do**
  foreach *public servant* i **do**
    determine contribution C_i;
  foreach *node* i **do**
    update indicator I_i;
  foreach *node* i **do**
    adapt allocation P_i;

Without priors on the initial values of the endogenous variables, we suggest random assignments and Monte Carlo simulations. Appendix D.4 elaborates on the stability of simulations in a Monte Carlo setting. Finally, while Algorithm 1 considers three exogenous inputs, we should point out that there can be additional sources of information, for example, a vector specifying how much transformative resources in each policy issue are flexible enough to be reallocated. This is a typical in federations, where the sub-national governments may receive resources from the federal authority with 'strings attached'; *i.e.* they can only be allocated to the issue at stake. Of course, such kind of data may only be available to certain government authorities. Nevertheless, the possibility of incorporating it into the analysis of PPI speaks to its flexibility and usefulness to policymakers.

4 Data

The collaboration with the UNDP focused in Mexico because, unlike in this country, national and sub-national governments have produced data on values that they want to reach for different development indicators (the goals) and budgetary information that can be linked to the SDGs. In this paper, we concentrate on the federal case. In this section, we present all these data, leaving further details about their pre-processing and normalization procedures for Appendix A.

4.1 Development indicators

We compiled a dataset with 141 national-level development indicators from Mexico, covering the period 2006-2016. Each SDG is covered by at least one indicator. Given the current social context of Mexico and the interest of the stakeholders of the project, special attention was given to collecting indicators from SDG 16. In fact, we split SDG 16 into its two components: *peace and justice* (SDG 16a) and *strong institutions* (SDG 16b). This is separation is important in the Mexican context as the former covers violence issues while the latter touches on anti-corruption policies. All indicators have been pre-processed so that their values are in the range [0,1], and larger magnitudes denote better outcomes.

A limitation of the official UN SDG database is that many indicators lack comprehensive time coverage. For this reason, we have collected data from additional sources and performed a manual classification into the SDGs. Finally, we labeled each indicator as *instrumental* or *collateral* according to the inputs received in a stakeholder workshop co-organized by the UNDP and the Mexican National Laboratory for Public Policy. Figure 2 shows the total number of instrumental and collateral indicators by SDG. Clearly, there are enough instrumental topics to define policy priorities across all the SDGs.

Figure 2: Number of indicators by type and SDG

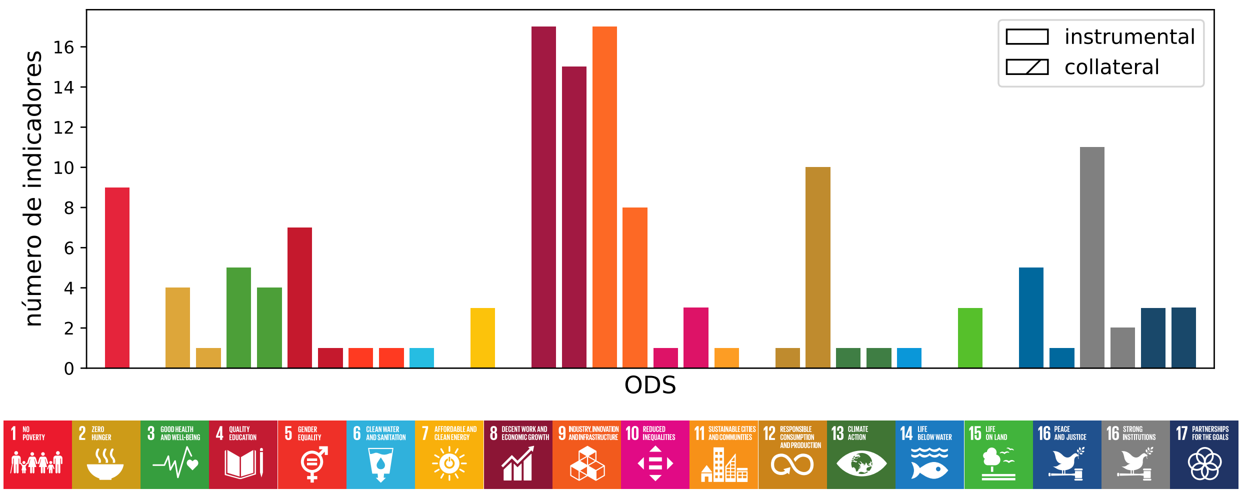


Table 1 shows summary statistics of the data by source.[8] The most salient feature in this table is the relative large number of missing observations coming from CONEVAL (the watchdog of social policy in Mexico). This is because CONEVAL's indicators are constructed on a bi-annual basis. Overall, whenever we encounter missing values, we impute them through linear interpolations. Further details can be found in Appendix A.[9]

[8]We provide the complete list of the indicators in Appendix C.

[9]We performed two particular data manipulations in order to include indicators that were of high interest to the stakeholders of the project. First, CONEVAL's indicators start in 2008 and are collected on the previous year to their publication. In order to make these data compatible with our sample, we lagged them by two years, meaning that the observations for 2008 correspond to those from 2006 in our data. Knowingly of the statistical shortcomings of such measure, this decision was concerted with the stakeholders, who dimmed important to have multiple indicators on poverty. In the end, remember that the analysis is rather a proof of concept, not a formal evaluation with definite estimations. The second adjustment comes from the fact that the index of perception of corruption by Transparency International had a methodological change

Table 1: Development-indicator data by source

Source	Indicators	SDGs	Observations	Missing	Mean	Min	Max
CONEVAL	8	1	88	40	0.75	0.35	0.95
FAO	3	2	33	0	0.51	0.14	0.92
INEGI	1	1	11	0	0.41	0.4	0.43
Observatory of Economic Complexity	1	1	11	0	0.71	0.67	0.73
The Conference Board	2	1	22	0	0.6	0.56	0.62
Transparency International	1	1	11	0	0.3	0.27	0.33
UN	32	14	352	68	0.58	0.0	1.0
WDI	1	1	11	0	0.0	0.0	0.0
World Bank	29	11	319	22	0.54	0.0	1.0
World Economic Forum	62	8	682	0	0.46	0.04	1.0
World Travel & Tourism Council	1	1	11	0	0.15	0.14	0.16

CONEVAL: Mexican institution responsible for the evaluation of social social policy.
FAO: Food and Agriculture Organization of the United Nations.
INEGI: Mexcian national statistics bureau.
UN: United Nations.
WDI: World Development Indicators.

4.2 Governance

For the governance parameters related to the quality of monitoring efforts φ and the strength of the rule of law τ, we use data from the Worldwide Governance Indicators. These indicators reflect the perception of citizens, entrepreneurs and experts in the public, private and NGO sectors. Although perception-based indices have well-known limitations, they are still one of the best metrics used in corruption studies. The indicator of *control of corruption* reflects the quality of the monitoring efforts by the central authority, which is an important element in the model. The indicator of *rule of law*, on the other hand, captures the quality of institutions designed to reassure a law-abiding society. Just like with the SDG data, we normalized these indicators using a worldwide sample, so their values fall within [0,1], reflecting the state of public governance in Mexico with respect of the rest of the world.

in 2012. This change increased the level of the indicator by one order of magnitude. Thus, we divided all observations after 2011 by 10.

4.3 Network

The network of conditional dependencies between development indicators can be obtained through various methods, each one implying certain assumptions about the data and about its underlying generating mechanisms. Ospina-Forero et al. (2019) provide a comprehensive survey on this topic. We must highlight that estimating large networks from time series is a burgeoning research area, so there is no strongly preferred method or accepted gold standard, especially given that development indicators tend to be quite short; with generally no more than 10 observations. Our method of choice is the *Sparse Gaussian Bayesian Networks* approach, developed by Aragam et al. (2018) and accessible through the `R` package `sparsebn`. We employ `sparsebn` to estimate the conditional dependencies network of Mexico using the time series constructed from its development indicators. The method estimates a structural equation model and returns a weighted directed network of conditional dependencies where the edges have been filtered in order minimize potential overfitting (hence the sparseness of the topology).

We represent the network as a matrix $\mathbb{A}$ where the dependencies or edges go from rows to columns. Hence the strength of a dependency $i \to j.$ is indicated by the weight $\mathbb{A}_{i,j}$. If the sign of the weight is positive(negative), we have a synergy(trade-off). $\mathbb{A}$ should be interpreted as a stylized fact –a collection of conditional dependencies– that the model takes into account to explain the dynamics of the indicators. Table 2 provides descriptive statistics of the estimated network.

4.4 Goals

When calibrating the model to historical data, it is necessary to assume that the government's aspirations or goals are the final values of the indicators. However, from a prospective point of view, the aspirations may be any hypothetical combination of values for the indicators (higher than their initial values). In Mexico, and in many Latin American countries, every government has to produce a development plan document

Table 2: Network statistics

SDG	In(Out) Degree Synergies	In(Out) Degree Trade-offs	In(Out) Strength Synergies	In(Out) Strength Trade-offs
1	2.44 (2.78)	2.33 (2.67)	1.79 (2.26)	1.43 (2.49)
2	1.4 (3.8)	1.8 (2.6)	1.65 (3.48)	0.53 (2.76)
3	1.22 (0.56)	1.22 (1.33)	1.33 (0.78)	1.53 (2.91)
4	2.0 (2.38)	1.62 (2.0)	1.39 (1.34)	1.15 (1.64)
5	3.5 (1.5)	0.5 (3.0)	2.12 (0.23)	0.15 (0.35)
6	1.0 (0.0)	0.0 (0.0)	0.01 (0.0)	0.0 (0.0)
7	2.67 (0.67)	1.67 (2.33)	1.17 (0.43)	0.09 (1.44)
8	2.5 (3.5)	1.72 (2.06)	1.7 (3.05)	1.31 (0.85)
9	2.76 (2.52)	1.68 (1.56)	2.16 (1.45)	1.56 (0.44)
10	2.0 (3.25)	2.25 (3.0)	2.32 (1.21)	0.85 (1.16)
11	1.0 (2.0)	3.0 (1.0)	0.0 (10.38)	0.0 (5.19)
12	2.45 (1.27)	2.55 (0.82)	0.5 (0.39)	1.35 (2.09)
13	2.5 (4.0)	2.0 (3.5)	3.82 (5.03)	1.54 (2.82)
14	2.0 (3.0)	0.0 (0.0)	5.92 (0.08)	0.0 (0.0)
15	2.0 (1.0)	1.33 (1.33)	1.94 (0.7)	0.05 (0.25)
16a	1.83 (1.5)	2.67 (1.0)	1.16 (0.64)	2.22 (1.43)
16b	2.77 (2.0)	2.38 (1.85)	2.08 (0.62)	1.32 (0.37)
17	3.5 (2.0)	1.33 (2.33)	1.05 (0.54)	1.18 (1.28)

The network has been divided into synergies and trade-offs. Due to the uneven distribution of indicators across SDGs, we normalized the statistics by the number of indicators in the relevant SDG. In(out)-degree is the number of incoming(outgoing) connections to(from) a node. The in(out)-strength is the sum of the weights of all incoming(outgoing) connections to(from) a node.

at the beginning of its term. The purpose of such document is to provide clarity on its development strategy and its objectives, facilitating the evaluation of its progress. For the federal Mexican government, this document is the National Development Plan (NDP) and, in its 2019-2024 edition, it consists of 234 objectives, 67 of which have concrete development indicators. That is, this document present the goals that the government –taking office in 2019– wants to achieve in its six-year term. Each of these indicators has been assign a development goal, *i.e.* a specific value to be achieved for the indicator. Thus, we use this information to conduct the prospective analysis in this paper. Figure 3 shows extracts of the Annex XVIII-Bis of the PND (Cámara de Diputados, 2019) containing some of its development goals.

In spite of providing explicit objective for the NDP indicators, these data do not map directly into the SDG development indicators. This is partly because governments evaluate their development through metrics that are specific to their context and needs,

Figure 3: Example

Para dar seguimiento al objetivo planteado, se propone el siguiente indicador:

Indicador 3.10.1: Intensidad de carbono por la quema de combustibles fósiles.

Descripción: emisiones de dióxido de carbono (CO2) por las actividades de quema de combustibles fósiles por dólar por paridad de poder de compra (PIB PPP).

Línea base (2016): 0.21 kilogramos de CO2 por quema de combustibles fósiles por dólar por paridad de poder de compra.

Meta 2024: 0.13

Para dar seguimiento y medir el avance en el logro del objetivo planteado se proponen los siguientes indicadores:

Indicador 2.4.1: Carencia por acceso a los servicios de salud.

Descripción: Mide la proporción de la población con carencia por acceso a los servicios de salud.

Línea base (2016): 15.5%

Meta 2024: 11.5%

Indicador 3.5.2: Índice de independencia energética.

Descripción: Es la producción nacional de energía primaria como proporción del consumo nacional de energía.

Línea base (2018): 0.7

Meta 2024: 1.0

Fuente: Sener

Indicador 2.11.2: Tasa de informalidad laboral (TIL-1).

Descripción: Mide el seguimiento de la formalización del empleo, para evaluar el mejoramiento de la calidad de los trabajos.

Línea base (2018): 56.58%

Meta 2024: 54.67%

Extracts from the Annex XVIII-Bis of Mexico's National Development Plan. Each box describes an indicator that will be used to evaluate progress in a specific policy issue of the NDP, as well as its baseline value (*Línea base*) and its target/goal (*Meta*). From left to right, the indicators track the following policy issues: carbon emissions from burning fuels; poor access to health services; energetic independence; informal labor. In the same order, the extracts were obtained from pages 187, 103, 166 and 128.

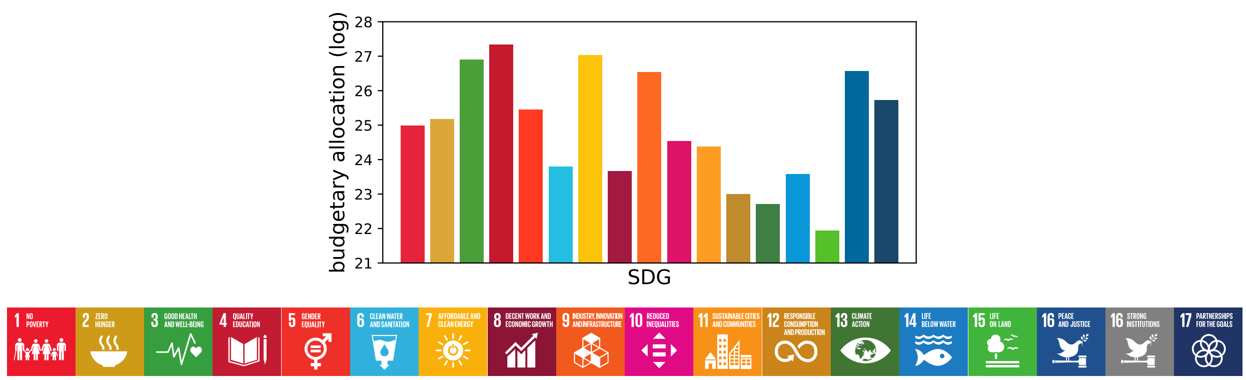
meaning that these indicators may not necessarily exist for other countries and, hence, are not part of official international SDG datasets. Nevertheless, the Mexican Treasury has developed a methodology to classify the NDP objectives into the SDGs (SHCP, 2017). (Cámara de Diputados, 2019, pp.216-219) provides the resulting classification of using such methodology in the 67 indicators of the NDP. Thus, in order to determine the development goals for the prospective analysis, we used the following procedure.

1. Compute the proposed growth rate for each indicator.
2. If an indicator in the SDG database corresponds to one in the NDP, assign the associated growth rate.
3. If an indicator in the SDG database has no corresponding indicator in the NDP, assign the average growth rate of the NDP indicators that are in the same SDG.

4.5 SDG budgeting

An additional innovation of the Treasury's methodology consists of linking expenditure categories into SDGs (SHCP, 2017). This is something unique to Mexico, and only now it is being emulated by other nations, so our study provides an early view to cutting-edge data and provides an innovative framework how to exploit it. The SDG-fiscal data consist of more than 1,500 expenditure records, each one categorized into one SDG. Figure shows the distribution of resources allocated through the Mexican budget for 2019 across the 17 SDGs. The fiscal data that has been classified into SDGs do not represent the entire budget of the government because some types of expenditure do not fit any SDG. In addition, these data does not distinguish between transformative and non-transformative resources. Thus, the exercise to be performed with this information is purely exploratory. Nonetheless, we consider important to demonstrate how PPI can exploit expenditure data. Finally, and to avoid any confusions, let us clarify that this budgeting data is only used as a component of the prospective analysis, it is not used anywhere else in the paper.

Figure 4: 2019 budget distribution across SDGs



The units in the vertical axis are current Mexican pesos in logarithmic scale.

5 Inferring historical policy priorities

For the retrospective analysis we use the development-indicator data, the governance parameters and the network; no target or budgetary data is employed in this analysis. In other words, PPI's three inputs are: (1) the initial values of the indicators, (2) the network of conditional dependencies, and (3) the final values of the development indicators. The inferred allocation profile P that reflects the policy priorities is estimated through the model since it is an endogenous variable generated by the government agent; it is not empirically observable. The first step for using PPI is calibrating the growth factors $\alpha_1, \dots, \alpha_N$ so that certain aspects of the simulated development indicators match their empirical counterparts (see section 3.3 and D for further details). There is a subtle but important implication in this calibration exercise: it is a retrospective inference the government's policy priorities during the sampling period (according to PPI's underlying theory). Furthermore (and a significant improvement over the CCG model), by finding the growth factors that make all indicators reach their final values in unison, this retrospective estimation acquires temporal value, so the prospective analysis can be interpreted in calendar time.

5.1 Retrospective policy priorities

An estimated allocation profile consist of the average distribution of resources across independent Monte Carlo simulations. For a single simulation, we obtain n synthetic time series of how the government agent allocated resources $P_{i,t}$ across each indicator i and every period t. The inferred allocation profile is built by, first, computing the inter-temporal sum of allocations for each indicator. Then, by performing M Monte Carlo independent simulations, we obtain M vectors $P_{i,\cdot}$ of inter-temporal sums. Finally, for each indicator i, we compute the mean allocation across Monte Carlo simulations. Altogether, the inferred allocation to the instrumental indicator i is given by

$$\mathscr{P}_i = \frac{1}{M} \sum_{m}^{M} \frac{1}{L_m} \sum_{t}^{L_m} P_{i,t,m}, \tag{16}$$

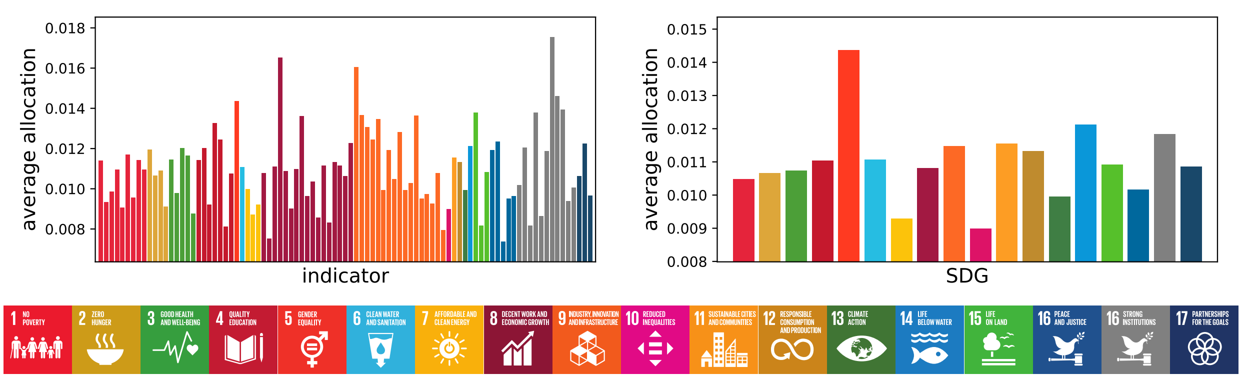
where L_m is the number of machine periods that the m^{th} simulation took to converge.

Figure 5 shows the inferred policy priorities that the Mexican government established during the sampling period.[10] According to this exercise, the most prioritized indicators tend to be in SDGs 8, 9 and 16b were the most prioritized. A more disag-

[10] Appendix F shows that these priorities are not trivial calculations. In other words, that the model (and its theory) provides new information that is not observable in development-indicator data.

gregate analysis can be found in Tables 3 and 4, which show the 10 most and least prioritized indicators respectively. There, we can see that the retrospective estimation suggests that the most prioritized issue was building statistical capacity, while the least one was fighting the organized crime. Both cases are consistent with the current situation of Mexico. On one hand, it is well known that recent governments have invested significant resources in creating regulatory organizations to guarantee the generation, transparency and protection of public data. This is an important part of the statistical capacity of the state. On the other hand, the spike in violence and crime across the country in the last decade speaks to the low priority that this policy issue has received in recent administrations.

Figure 5: Retrospective allocation profile



Each bar in the left panel can be interpreted as a share of the total amount of transformative resources, so they add up to one. The bars on the right panel are averages of the bars in the left panel, computed for each color.

Table 3: Most prioritized indicators

SDG	Indicator
16b	Overall level of statistical capacity
8	Burden of customs procedures
9	Quality of overall infrastructure
16b	Legal rights index
5	Proportion of seats held by women in national parliaments
16b	Political stability and absence of violence/terrorism: estimate
15	Average proportion of terrestrial key biodiversity areas
16b	Intellectual property protection
9	Capacity for innovation
9	Quality of roads

The indicators have been ordered from top to bottom in ascending order of priority. In other words, the first item is the most prioritized.

Table 4: Least prioritized indicators

SDG	Indicator
16a	Organized crime
8	Efficiency of government spending
9	Investment in transport with private participation
4	Quality of management schools
15	Average proportion of mountain key biodiversity areas
16b	Property rights
8	Tax revenue
8	Cooperation in labor-employer relations
16b	Judicial independence
3	Proportion of the target population with access to measles-containing-vaccine second-dose

The indicators have been ordered from top to bottom in ascending order of priority. In other words, the first item is the least prioritized.

6 Prospective analysis

We dedicate most of the analysis to prospective exercises because, from our interactions with policymakers, this type of inference is most useful to the stakeholders. Here, we employ the data on targets obtained from the NDP. Hence, the three inputs for PPI's prospective analysis are: (1) the final values of the development indicators, (2) the network of conditional dependencies, and (3) the goals constructed from the NDP data.

The inferred prospective allocation profile represents the policy priorities that the Mexican government would establish if it would truly pursue the goals presented in the

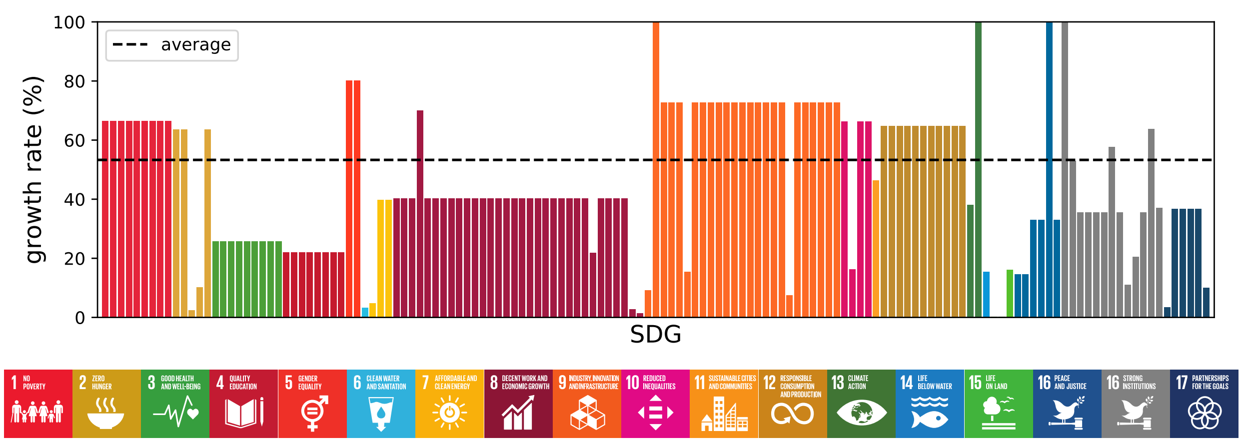
NDP. This, like with any prospective analysis, comes with a couple of assumptions. First, it is assumed that the growth factors $\alpha_1, \ldots, \alpha_N$ are the same as the ones obtained from the retrospective analysis. For the type of short-term studies that PPI is design to support, this assumption is quite reasonable because the growth factors capture structural features whose changes take place in the long run. Second, the network of conditional dependencies is supposed to be the same as the one used in the retrospective analysis. The argument here is similar to the one for the growth factors: the network topology and the weights of its edges represent long-term structural relations. Note that assumptions 1 and 2 do not prevent the user from performing experiments modifying the growth factors or the network change. For instance, an interesting exercise would be to understand how promoting certain synergies (strengthening positive edges) and discouraging certain trade-offs (removing negative edges) can affect the feasibility of a development strategy. This type of experiments has been considered in Figure 1 under interventions that fall within structural reforms. Third, as with the retrospective analysis, allowing P to be fully endogenous assumes fluidity in the allocation of transformative resources. This is a reasonable assumption given that most fiscal rigidities take place in committed resources that are not transformative. Nevertheless, should the user have data on the fiscal rigidities of transformative resources, it can also be used in PPI (as a fourth input), something that we explore in section 6.5 with the budgetary data.

We organize the prospective analysis in this section in the following way. First, we present the development goals constructed from the NDP. Second, we obtain the policy priorities that Mexico would establish to reach these goals. We call this inference *fluid priorities* because the allocation profile P is completely endogenous. Third, we assess the feasibility of the fluid priorities by estimating the years it would take the government to achieve the proposed goals. Fourth, we compare the budgetary data with the fluid priorities. This allows us to measure the level of policy coherence through the index created by Guerrero and Castañeda (ming). Fifth, explore *rigid priorities* by fixing P as the budgetary data, and evaluate their feasibility. Sixth, we introduce a methodology for discovering accelerators through PPI and compare our results against the naïve approach of using network-connectivity metrics. Seventh and last, we discuss on how PPI has been validated and how stringer validation procedures can be developed with the availability of more and better data.

6.1 Development goals

Figure 6 presents the development goals of the PND mapped into the SDG indicators of our database. On average, the proposed growth rate for the indicators is higher than 50%. SDG 16b presents more heterogeneity in goals because it has more direct matches between PND and SDG indicators. In contrast, the bars that have identical values are the result of assigning the average growth rate of all indicators within the same SDG (because the indicator at stake did not have a direct match with an indicator from the NDP; see section 4.4 for more details). We build the prospective target vector T with this information by taking the final values of the indicators and growing them according to the associated rates.

Figure 6: Prospective development goals

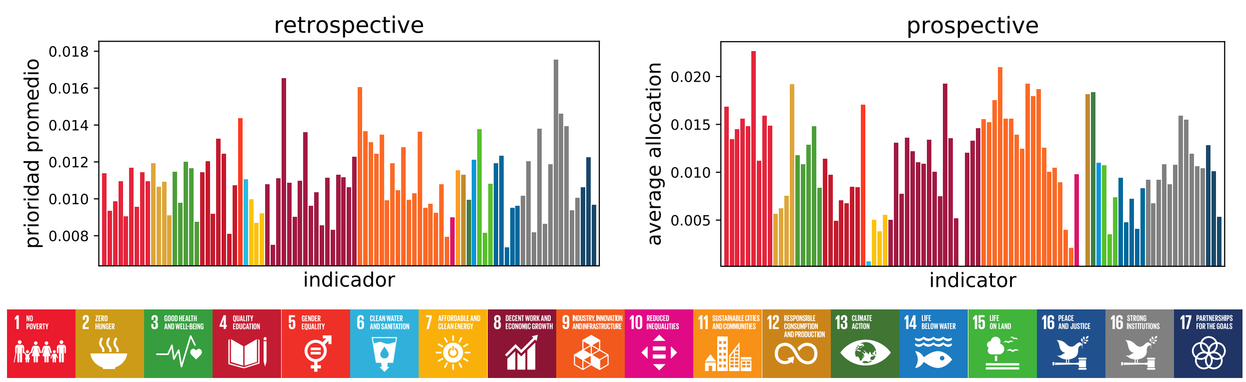


6.2 Fluid policy priorities

Figure 7 presents the result of this estimation, and compares it with the retrospective policy priorities. The prospective allocation profile should be interpreted as the policy priorities that the Mexican government would establish –under the assumptions previously discussed– if it would truly pursue the proposed goals. Note that, since the goals are based on the NDP, discrepancies between the prospective and the retrospective profiles can be interpreted as: *how different are the government's actions with respect to the previous administrations?*. For this reason, Figure 7 presents both the prospective and the retrospective estimations.

From comparing both allocation profiles, we can see important increments in SDG 1, 3, 9, 10, 12 and 13. On the other hand, the SDGs with a noticeable decrease in priorities are 4, 6, 7 and 16b. Whether the inferred prospective allocations are the actual actions that the government will reflect in its budget is a different story, one that we explore in section 6.4. The important takeaway from this exercise is that PPI can be used to assess how different would be the actions required to achieve a new set of goals, given what previous administrations have been doing; an important piece of information to asses the feasibility of a development strategy.

Figure 7: Prospective policy priorities under perfect fluidity



Each bar in the left panel can be interpreted as a share of the total amount of transformative resources, so they add up to one. The left panel corresponds to the retrospective policy priorities presented in Figure 5. The right panel presents the fluid policy priorities inferred from the prospective analysis.

6.3 Feasibility under fluid priorities

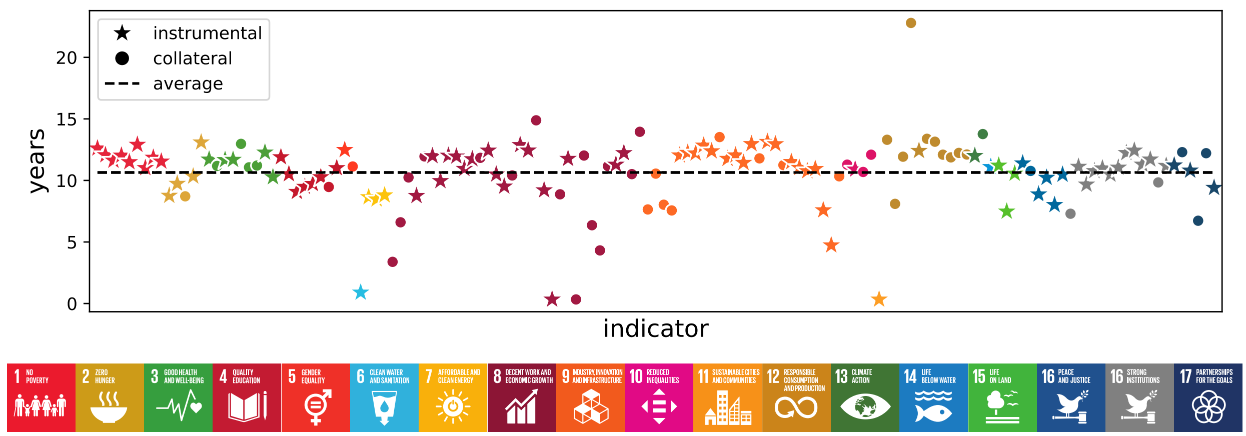
As we have previously mentioned, the calibration procedure gives PPI temporal value in the sense that machine ticks can be mapped into calendar time. Consequently, it is possible to estimate the calendar time that it would take to reach the development goals of the NDP. Thus, by measuring convergence time, we can assess the feasibility of reaching the proposed development goals. For example, in the case of Mexico, a goal that takes more than 6 years to be attained requires more than one administration, while one that takes more than 30 years can be considered unfeasible. We perform this exercise and present the convergence times in Figure 8.

Before interpreting the results, it is important to point out a caveat in this analysis and its potential solutions. Because there is not a one-to-one correspondence between the SDG indicators and the NDP, the way in which the target vector was constructed involved computing growth rates from the NDP and mapping them into the SDG indicators. This introduces a bias in those indicators where Mexico lags the most. For instance, if an SDG indicator has a value near zero, it means that, across all the countries in the dataset, Mexico is among the worst performers in this topic. When multiplying that value by $(1+growthRate)$, one would obtain a very low goal, so a minuscule gap to be covered. In reality, the government would not have such a low goal for that indicator, but due to the lack of information regarding this goal, its approximation through growth rates is the second best. The consequence is that those indicators in which Mexico performs extremely bad will likely show fast convergence times. In order to fix this, it is important to elicit development goals in the original units of the indicators, and this can be done by either working closely with the stakeholders, or by selecting indicators for which hypothetical goals could be easily established (*e.g.*, through an analysis using only NDP indicators rather than SDG ones).[11] The main implication of this bias is that it underestimates the average convergence time in the prospective analysis. Thus, the interpretation in terms of feasibility might be overoptimistic.

[11]Due to the SDG–oriented and method-development nature of the project, it was not possible to elicit SDG development goals of the federal government, other than those published in official documents.

Figure 8 shows the convergence time of each indicator in years. On average, it would take more than 10 years to reach the goals proposed in the NDP. That is, it would take two administrations to get there. One indicator in SDG 12 stands out as the most difficult to achieve: *ethical behavior of firms*, with almost 25 years of convergence time. Overall, these results suggest that the NDP demands a multi-term strategy in order to be considered feasible.

Figure 8: Convergence time under fluid priorities



6.4 Budget and policy coherence

Recall that, according to PPI's theory, if the government truly wants to achieve the goals proposed in the NDP, it would establish the prospective priorities estimated in Figure 7. We also mentioned that, whether this is reflected in the government's actions or not is a different questions, one that we consider here. Let us consider the budgetary data presented in section 4.5, and assume that it reflects the policy priorities that the government intends to establish in the next six years.[12] If the budget proposed by the federal government would be coherent with the NDP, then it

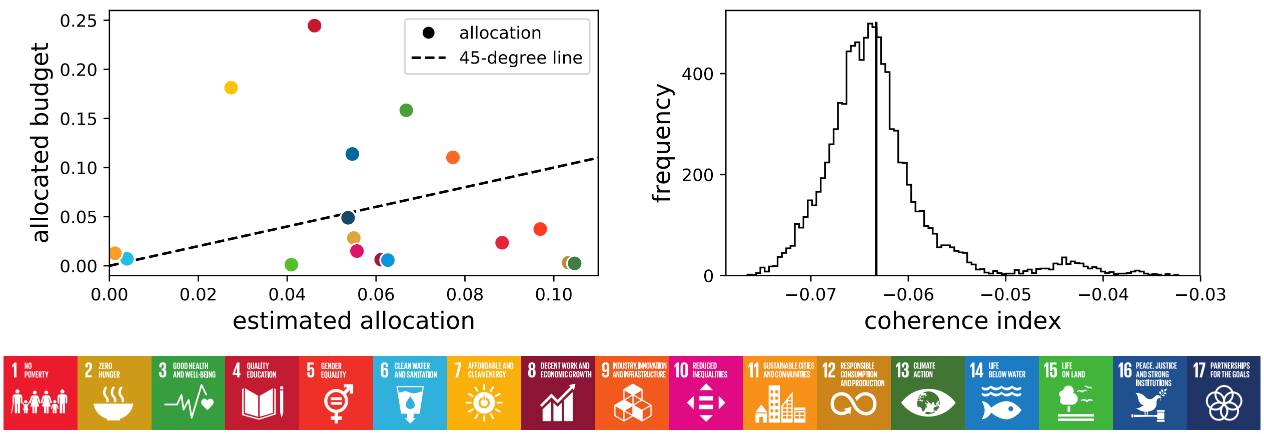
[12]Of course, this data does not differentiate committed from transformative resources. However, it is useful to show this exercise to demonstrate how, should such fiscal data become available, PPI can be enriched.

should resemble the prospective policy priorities. This is so because the prospective priorities are, effectively, a counterfactual of what the government would do under the NDP. Comparing this counterfactual to the fiscal data allows measuring the level of coherence of the budget through an index developed by Guerrero and Castañeda (ming).

Figure 9 shows the result of this exercise. The left panel presents the discrepancies between the budget and the priorities, both at the SDG level because the budgetary data is linked at an aggregate level, not at the indicator level. Let us concentrate on the right-most dot in the panel, which corresponds to SDG 13. The horizontal axis suggests that government would establish an average allocation of more than 10% to SDG 13. However, the vertical a axis shows that the actual budget that was approved for such purpose is of less than 1%. Clearly, the government is under-spending in *climate action.* Under- and over- expenditures are also called an allocative inefficiencies and, together with technical inefficiencies, they are key to understand development (Izquierdo et al., 2018). The more allocative inefficiencies we find in the plot, the more likely that the coherence index will suggest incoherence between the federal budget and the NDP.

The coherence index behaves like a correlation coefficient: positive values indicate coherence, while negative ones incoherence. An index of one suggests full coherence, meaning that all the dots in the left panel would lie on the dotted line because both the budget and the prospective allocation are identical. The right panel shows the distribution of the coherence index (see Guerrero and Castañeda (ming) for details on the estimation procedure). Clearly, the budget is incoherent with the government's aspirations (at least at the SDG level). Nevertheless, the magnitudes of the indices are not to high, suggesting that moving to the positive side is feasible. Note that the coherence index is specific to the set of indicators, goals, growth factors and spillover network. Therefore, an incoherent budget may not necessarily mean that the priorities are wrong, it may suggest, instead, that the goals are unrealistic (probably the case in the goals established for SDG 13). In any case, the usefulness of PPI becomes evident because it allows assessing both sides of the same coin.

Figure 9: Budget and policy coherence



6.5 Rigid policy priorities

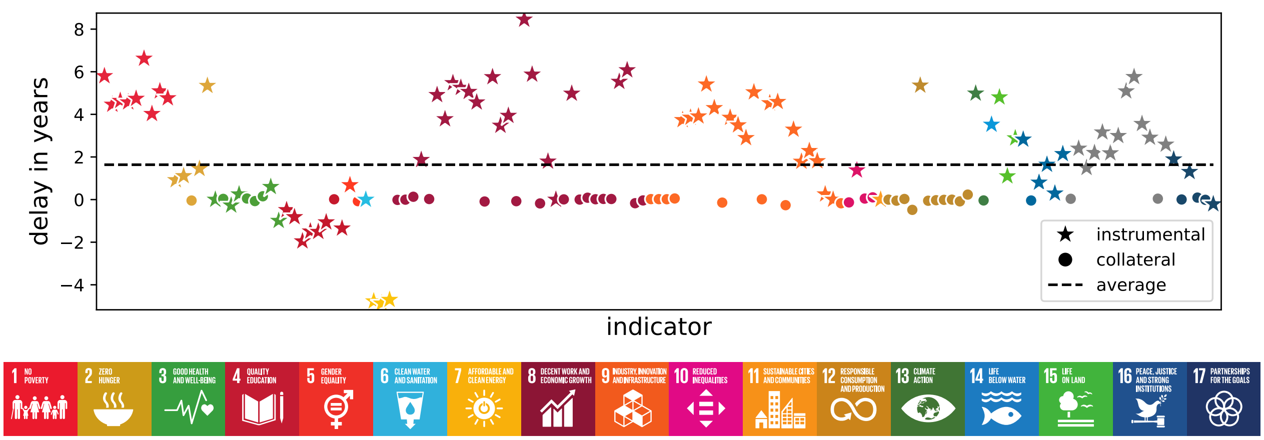
Suppose that the budgetary data reflects indeed the policy priorities, and that the government is not willing to change them throughout its administration. Here, we are in a context of fiscal rigidities. For PPI's model, it means that the government agent cannot adapt its allocations and, instead, that it needs to fix them to a given priority vector (in this case, the budget data). Then, the natural question that arises in this scenario is whether following the approved budget will make the NDP more or less feasible. We can assess this by comparing the convergence times under the budgetary data against the ones obtained from the prospective priorities. New sets of simulations need to be run for (1) taking the budgetary data as the given priorities[13], and (2) using the prospective priorities obtained in section 6.2 as the given ones. The reason why we cannot use the convergence times obtained in section 6.3 is because that exercise assumes fluid priorities, while the adoption of the budget assumes rigid ones. This difference in levels of rigidity induces differences in convergence times, as shown

[13]Since the budget data is at the SDG level, the indicator-level budget was constructed by assigning each indicator an equal fraction of the total allocation given to the associated SDG.

by Castañeda and Guerrero (2019). Therefore, the correct way to measure delays or savings in convergence time is by having the same level of fiscal rigidity.

Figure 10 show the the differene in convergence time between adopting the federal budget or the propsective allocations. On average, adopting the aproved budget would produce an average delay of nearly 2 years with respect to the adoption of the prospective priorities. In some cases, the delays can be of the order of 6 or 7 years. The indicator where the biggest delay takes place is *cooperation in labor-employer relations*, from SDG 8. We can also observe some savings, especially in the indicators related to SDG 7. Overall, this exercise shows how PPI can be combined with fiscal-SDG linked data, a very promising research agenda as more and better public expenditure data is becoming available every day.

Figure 10: Delays in reaching the development goals



6.6 Accelerators

The last of our prospective analysis touches on a topic that has become highly discussed in the SDG literature:*accelerators*. According to the UNDP's SDG Accelerator and Bottleneck Assessment (ABA) tool (gar, 2017, p. 6), an accelerator is a "*–development policy and/or programme areas– that will accelerate progress across the SDGs and the* *national development goals, and the corresponding drivers that enable their progress*". Accordingly, the purpose of identifying accelerators is to prioritize them in order to trigger positive multiplied effects across SDGs. In principle, this very notion appeals to the network of interdependencies. While the ABA tool takes a more qualitative approach, other groups have emphasized the use of SDG networks to prioritize topics with a high degree of connectivity (Weitz et al., 2018). In this section, we (1) demonstrate that the approach of using networks to identify accelerators is not only naïve but may be counterproductive, and (2) provide a more comprehensive way to identify accelerators by combining PPI with heuristic optimization methods.

What does an accelerator mean in PPI? Under PPI, an accelerator cannot be just an indicator with a high degree of connectivity for the following reasons:

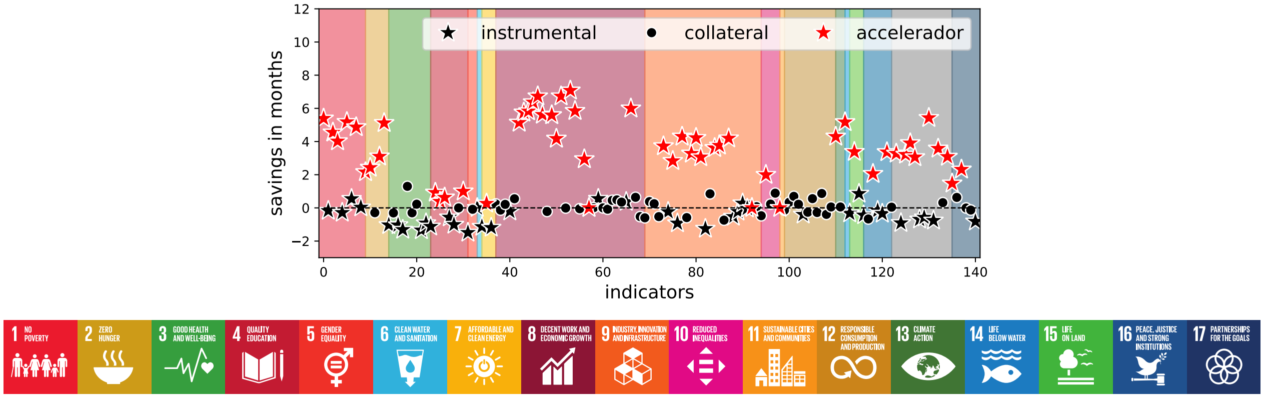
1. Only instrumental indicators can be accelerators because they are the only ones that can be intervened through policy.
2. The network is not causal, so it is erroneous to assume that one can manipulate the indicators. Instead, one tries to improve them through policy interventions.
3. Given that there is always a budgetary constrain, when one increases the priority in one topic, it takes away resources from other policy issues. Thus, an accelerator cannot be considered in isolation from those topics that are losing resources (and this is different from network dependencies).
4. An indicator with high connectivity may be highly inefficient, so allocating more resources without solving the associated public governance problems may render in ineffective policies.
5. Shifting priorities to a node may affect the incentive structure of the policymaking agents. For example, an agent may become inefficient because it receives more spillovers from the intervened policy issues.

Clearly, in the identification of accelerators, there is more to than the eye meets. Much of this complexity comes from the political economy underlying the policymaking process, and this is something that data, by itself, cannot reveal. Fortunately, the philosophy of PPI appeals to a generative approach to deal with such complexities, something that we exploit to shed new light on the identification of accelerators. Let us concentrate on the case where the policy priorities are given by the budgetary data. Suppose that the government takes 10% of the budget assigned to each indicator, and reallocates it to the accelerators. That is, we choose the accelerators ex ante and give them the 10% of the budget in equal portions. Then, we run PPI under full rigidity and obtain the convergence time, which is the evaluation criterion for this exercise (but one could perfectly use other criteria). Then, the question is: which of the 91 instrumental indicators should be chosen as the accelerators?

Let us begin with the naïve approach, where the accelerators should be those indicators whose total outgoing weights in the network are positive. That is, these are nodes with a net positive conditioning on other nodes, possibly because they have a high degree of positive outgoing links, but also because their positive outgoing edges outweigh their negative ones. We reallocate 10% of the budget to these indicators and run PPI. Then, we compute the difference in convergence speed with respect to the times obtained in section 6.5 for the budget priorities. Figure 11 shows the result. Overall, the naïve approach identifies 54 accelerators, and generates an average saving of 1.29 months, with individual savings per indicator of up to 8 months. Note, however, how some accelerators generate zero or near-zero savings. This means that the political economy process prevents them from being real accelerators, in spite of having net positive outgoing spillovers.

If the naïve approach to accelerators has clear limitations, how can we find the set of instrumental nodes that can best catalyze development by assigning them the 10% of the budget? A brute force strategy is to simply enlist all the possible sets of accelerators and run PPI for each one to select the best-performing set. The problem is that, given the high-dimensional nature of development, exploring all the possible sets becomes unfeasible very soon, for example, in this particular study, the number of all possible accelerator sets is $2^9 1$ (that is an order of magnitude of 27). In addition, the

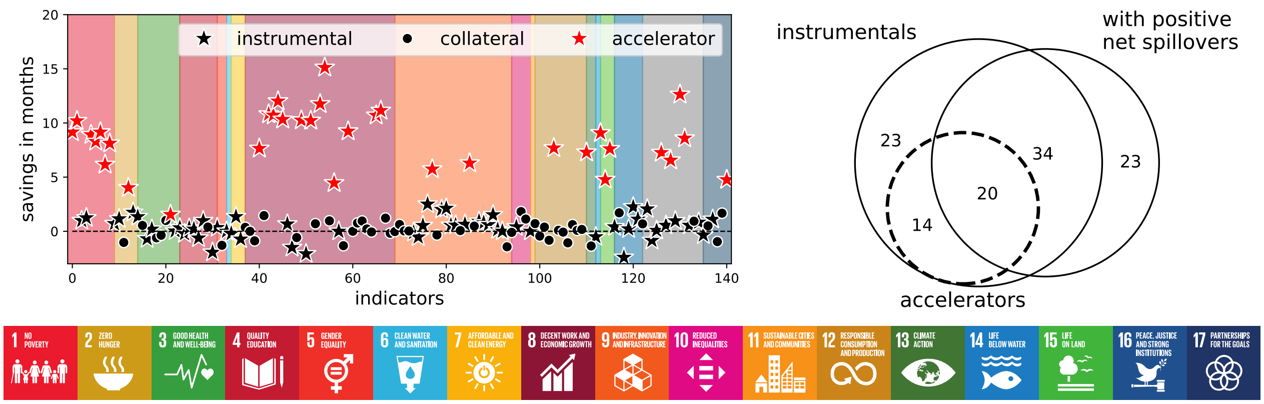
Figure 11: Accelerators identified through network connectivity



‘fitness landscape’ of this problem is far from beind smooth, so traditional optimization methods are bound to fail. Fortunately, heuristic optimization techniques can have been designed to deal with this problem. In this application, we employ one that stems from Complexity Science: *genetic algorithms* (Holland, 1975). The left panel in Figure 12 shows the results of our search for accelerator. In total, the genetic algorithm returned a set with 34 accelerators, considerably less than in the naïve approach. In this set, not all the SDGs have accelerators, for example, SDGs 4 to 7. Overall, allocating the 10% of the budget to these indicators generates an average saving of 2.25 months, twice as much as the naïve accelerators. Note that all the accelerators lie above the dotted line, and that, in some cases, there are savings of more than one year.

Note that, in the right panel, we present a Venn diagram showing the distribution of accelerators across the different types of indicators. Interestingly, 14 out of the 34 accelerators, have negative net spillovers, the exact opposite intuition to the naïve approach. This demonstrates how neglecting the political economy process can produce misleading policy prescriptions. Clearly, the identification of accelerators demands a systemic and complexity point of view; that is the very essence of the 2030 Agenda. This exercise shows the versatility of PPI and how, when combined with other methods from Complexity Science, it can be used to tackle challenging problems. A more thorough study could explore other heuristic optimization methods, as well are diverse reallocation rules and levels of fiscal rigidity. Here, we have opened the path to explore these possibilities.

Figure 12: Accelerators identified through PPI and heuristic optimization



6.7 A comment on validation

Given the policy relevance of PPI, a natural question that arises is how has it been validated. To answer this, it is important to first acknowledge that the meaning of validation varies across fields and methodologies. For this reason, we would like to comment on different ways in which we have validated PPI, and to discuss potential directions in which validation can be strengthened in the future as more and better data becomes available. Much of the validation has already been done in the original work of the CCG model (Castañeda et al., 2018), on which PPI is built. Hence, we discuss some of these validations and elaborate on new ones that were done for this paper.

First, let us discuss external validation in agent-based modeling, which typically means replicating an aggregate stylized fact by generating it from bottom-up, *i.e.* without assuming the mechanisms that will trivially lead to the aggregate result. In the early work on the CCG model, such type of validation was provided by replicating the cross-country distribution of corruption. By using an independent indicator on the diversion of public funds, the CCG model was able to generate endogenous levels of corruption that matched multiple features of this indicator: (1) a negative correlation between corruption and economic performance, (2) a large variation of corruption among middle-income countries, (3) no observations of low-income countries with low levels of corruption, and (4) no overlaps in corruption levels between low-income and high-income countries. The reason why this is considered a validation is because the model was not specified to guarantee such stylized fact. For instance, while indeed the differences in governance parameters contribute to higher or lower levels of corruption, the spillover networks could have topologies that prevent this stylized fact from emerging, for example, because the amount of synergies would encourage more corruption. This argument also applies to the micro-level social mechanisms, which leads us to internal validation.

Internal validation is common in the agent-based modeling literature and it consists in showing that the bottom-up replication of a stylized fact is sensitive to the social and behavioral mechanisms specified in the model. In other words, through sensitivity analysis, internal validation aims to show that without a relevant theory, the model would not be able to produce the relevant stylized facts. In this respect, the CCG model has been shown to fail in reproducing the aforementioned stylized facts when removing the directed-learning component of the agents, or the behavioral heuristics nature of the government. Another way to look at internal validity is by evaluating if a meso level level property that is not observable but theoretically expected emerges, and if such emergence is sensitive to the its theoretical drivers. In the CCG model, such meso level property is a negative relation between positive incoming spillovers and the agents' contributions. Here, PPI's theory suggests that, given everything else equal, an agent that receives more positive spillovers has incentives to become more inefficient because, in the eyes of the central authority, he/she is a proficient agent. Again, this is not guaranteed to happen because the learning process is heuristic, and the system is full of uncertainties. Nevertheless, Castañeda et al. (2018) show that this correlation emerges, and that it disappears if one removes the network from the system.

In studying policy coherence with the CCG model, Guerrero and Castañeda (ming) provide a 'soft' validation exercise. This consists of estimating the index of policy coherence for countries that are known to have been coherent with emulating specific economies in the past, for example, Korea following Japan or Estonia adopting the Nordic development model. If the coherence index is consistent with this qualitative narrative of successfully emulations, it provides further evidence that the inferred policy priorities contains valid information. This is the case in both examples.

Now that we discussed soft validation, we can also talk about stakeholder validation, a popular approach in participatory modeling. Here, the stakeholders of a project can validate the a methodology by (1) contributing to refine the model or the data through feedbacks throughout the project and (2) by determining whether the results of the analysis 'make sense' to some degree. During the UNDP project, different stakeholders from the federal- and state-level governments took part of two workshops where the methodology, data and results were presented and discussed. The stakeholders took part of an exercise in which they had to classify the indicator database into instrumental or collateral and, then, the results where discussed to arrive to a consensus. Hence, the stakeholders provided early feedback in refining the data that led to the analysis. Then, once presented with the results, the stakeholders acknowledged their intuitive nature. For example, the low priority given to SDG 16a, *peace and justice*, (see Figure 5) by the past administrations made complete sense to the participants, while the low feasibility of reaching the goals proposed for SDG 12, *responsible production and consumption*, was acknowledged by expert stakeholders. Therefore, in this particular study, PPI also enjoys if stakeholder validation.

Finally, it is important to mention that as more and better data becomes available, further validation tests could be developed. For example, with historical budgetary data on transformative resources at the level of each indicator one should be able to assess how well the model can replicate the expenditure patterns through the endogenous policy priorities. In the worst case scenario, the government heuristic could be refined to minimize the discrepancies between P and the expenditure data. From a machine-learning point of view, better development data could also allow out-of-sample validation. For instance, with long time series of development indicators, it should be possible to assess whether the synthetic indicators can correctly match values that are out of the sample. Then, through cross-validation, the calibration procedure could be further refined to match more features of the indicators. These are avenues that are highly promising, as they suggest that PPI can only improve with the availability of data in the future.

7 Conclusions

The 2030 Agenda, through its acknowledgement of the complexity of development, poses significant challenges. Fortunately, Complexity Science and Computational Sciences have developed tools to tackle these problems. This paper introduces one such tool: Policy Priority Inference. PPI takes into account the complexities of the political economy underpinning the dynamics of development indicators, and considers important institutional factors that are key to the success or failure of policy interventions. We explain the methodology and present applications for the case of the Mexican federal government. In particular, we show how to assess the feasibility of development goals, the coherence of policy priorities, and the identification of development accelerators. Overall, our applications demonstrate how PPI is well equipped to deal with the complexity challenges of the 2030 agenda, and they pave the way for new ways of analysing development strategies. Although we develop a specific model for the analysis, PPI should be though as a philosophy in which development should be understood as a complex adaptive process. This means that as more and better data becomes available, computer hardware improves, and our knowledge on human behavior, institutions and social interactions grows, better models can be developed within this framework. Ultimately, the PPI philosophy tries to provide a flexible and open scheme with which governments, NGOs, consultants, academics and the civil society can work towards meeting the goals of the current this and future international development agendas.

References

(2017). SDG Accelerator and Bottleneck Assessment. Technical report, United Nations Development Programme, New York, NY.

Alkire, S. and Foster, J. (2011). Counting and Multidimensional Poverty Measurement. *Journal of Public Economics*, 95(7):476–487.

Aragam, B., Gu, J., and Zhou, Q. (2018). Learning Large-Scale Bayesian Networks with the Sparsebn Package. *Journal of Statistical Software*, forthcoming.

Bank, W. (2017). *World Development Report 2017: Governance and the Law*. International Bank for Reconstruction and Development / The World Bank, Washington, D.C. OCLC: 956624957.

Baum, A., Hackney, C., Medas, P., and Sy, M. (2019). Governance and State-Owned Enterprises: How Costly is Corruption? IMF Working Paper 19/253, International Monetary Fund, Washington, D.C.

Cámara de Diputados (30-April-2019). Anexo XVIII-Bis.

Casini, L. and Manzo, G. (2016). Agent-based Models and Causality: A Methodological Appraisal. *The IAS Working Paper Series*, 2016(7).

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

Castañeda, G. and Guerrero, O. (2018). The Resilience of Public Policies in Economic Development. *Complexity*, 2018.

Castañeda, G. and Guerrero, O. (2019). The Importance of Social and Government Learning in Ex Ante Policy Evaluation. *Journal of Policy Modeling*.

Chambers, R. and Institute of Development Studies (Brighton, E. (2007). Poverty research: Methodologies, mindsets and multidimensionality. Working Paper, Institute of Development Studies at the University of Sussex, Brighton. OCLC: 225416493.

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford Univeristy Press, Oxford.

Guerrero, O. and Castañeda, G. (2019). Does Better Governance Guarantee Less Corruption? Evidence of Loss in Effectiveness of the Rule of Law. *arXiv preprint arXiv:1902.00428*.

Guerrero, O. and Castañeda, G. (forthcoming). Quantifying the Coherence of Development Policy Priorities. *Development Policy Review*.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press.

IMF (2019). Fiscal Monitor: Curbing Corruption. Technical report, International Monetary Fund, Washington, D.C.

Izquierdo, A., Pessino, C., and Vuletin, G., editors (2018). *Better Spending for Better Lives: How Latin America and the Caribbean Can Do More with Less*. Inter-American Development Bank.

Klitgaard, R. (1988). *Controlling Corruption*. University of California Press.

McGillivray, M. and Shorrocks, A. (2005). Inequality and Multidimensional Well-Being. *Review of Income and Wealth*, 51(2):193–199.

Ospina-Forero, L., Castañeda Ramos, G., and Guerrero, O. (2019). Estimating Networks of Sustainable Development Goals. *Working Paper*.

Persson, A., Rothstein, B., and Teorell, J. (2013). Why Anticorruption Reforms Fail—Systemic Corruption as a Collective Action Problem. *Governance*, 26(3):449–471.

Rose-Ackerman, S. (1975). The Economics of Corruption. *Journal of Public Economics*, 4(2):187–203.

SHCP (2017). Vinculación del Presupuesto a los Objetivos de Desarrollo Sostenible. Anexo 2 de los Lineamientos para el Proceso de Programación y Presupuestación para el Ejercicio Fiscal 2018, Secretaría de Hacienda y Crédito Público, CDMX.

UN General Assembly (2015). Transforming our world : The 2030 Agenda for Sustainable Development. Technical Report A/RES/70/1.

Weitz, N., Carlsen, H., Nilsson, M., and Skånberg, K. (2018). Towards systemic and contextual priority setting for implementing the 2030 Agenda. *Sustainability Science*, 13(2):531–548.