QUANTIFYING THE COHERENCE OF DEVELOPMENT POLICY PRIORITIES

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

Structured Abstract

Motivation

Over the last 30 years, the concept of policy coherence for development has received attention among academics, practitioners and international organizations. However, its quantification and measurement remain elusive.

Research questions

To address this challenge, we develop a quantitative counterfactual-based framework to measure the coherence of policy priorities for development.

Approach of methods

Our procedure takes into account the country-specific constraints that governments face when trying to reach development goals. Hence, we present a new definition of policy coherence where resource allocations are employed as the baseline to construct an index.

Results

We validate this coherence index by producing country-specific estimates that are consistent with well-known cases of development strategies (both successful and failed).

Conclusions

Our results highlight the limitations of assessing coherence in terms of benchmark comparisons using development-indicator data. Altogether, our framework sheds new light in a promising direction to develop bespoke analytic tools to meet the 2030 Agenda.

Key words: computational; development; index; multidimensional; network; OECD; policy coherence; policy priorities

[1] The Alan Turing Institute, London and Department of Economics, University College London, London. Corresponding author. Email: oguerrero@turing.ac.uk

[2] Center for Research and Teaching in Economics (CIDE), Mexico City, Mexico.

1 INTRODUCTION

When governments aim to improve the performance of specific socioeconomic indicators, the policies they formulate and implement are often marked by inefficiencies of various kinds. In addressing policy design and implementation, academics and development consultants often highlight the importance of coherence, acknowledging the fact that development goals and policies are multidimensional. Nevertheless, the term "policy coherence" is a loosely defined concept that has different meanings across various researchers and organizations. The need for conceptual clarity and unambiguous measurements calls for a redefinition of the term. In this article, we develop such definition and construct a relevant metric. The proposed index makes it possible to estimate how coherent a country's policy priorities are when it attempts to reach multidimensional development goals.

Traditionally, assessing policy coherence involves qualitative methods such as analysing official speeches and documents that signal the degree of government policy alignment with a set of established goals (OECD, 2015, 2016, 2017, 2018; European Commission, 2019). While these approaches may give initial pointers, we still lack precise quantitative measures of policy coherence, because policy priorities are not directly observable in discourse or public expenditure data.

The construction of a metric that quantifies policy coherence is critical for several reasons. First, it provides a less discretionary way to measure the level of a government's commitment to reaching certain development goals. Second, it allows comparisons between countries and regions, which is extremely helpful in order to evaluate and rethink international development agendas. Third, a quantitative metric provides an alternative—or complementary—strategy to qualitative expert advice (which tends to be expensive). Fourth, it helps governments design timely responses by informing the reorganization of policy priorities. Fifth, it reinforces the need for evidence-based policy-making towards the United Nations 2030 Sustainable Development Goals (SDGs) Agenda.

Despite the potential benefits of a coherence metric, there are as yet no scalable and robust indices. This is due to various data-related and theoretical challenges that need to be overcome. Here, we mention a few. First, empirical data such as development indicators do not reflect policy priorities, but are rather the outcome of complex policy-making processes. Therefore, it is necessary to develop theoretically informed socioeconomic models that simulate such processes. We suggest that such models be computational. Second, in order to account for spill-overs between policy issues (important in the SDG literature, and also called 'synergies' and 'trade-offs'), new statistical methods are necessary. In particular, network-estimation methods are required. Third, when implementing a computational model, it is crucial to account for the institutional context of the country in question;

in particular, for mechanisms related to governance and corruption. Fourth, policy coherence is intimately related to the specific context of each country. Hence, if possible, rather than pooling cross-national data, coherence should be measured independently for each nation. The challenge lies in the fact that country-specific data on development indicators tend to be coarse-grained, so the associated datasets have a small number of observations. Clearly, there are numerous challenges in building a coherence index. Our work sheds new light on how to overcome some of them.

As explained above, our approach to policy coherence is built on policy priorities. These priorities, in turn, are reflected in the volume of resources (fiscal and otherwise) that the government allocates to transformative policies. However, public expenditure data do not reveal such information because they are typically coarse-grained, are not matched to specific development indicators, and do not differentiate between transformative and non-transformative resources (e.g. maintaining road infrastructure vs. expanding the highway network). In order to build a quantitative metric of policy coherence, it is therefore necessary to estimate policy priorities by using a socioeconomic model. The model should contain the essential mechanisms through which the government allocates resources, and through which such resources are transformed into development outcomes. The idea is to use such model to simulate development indicators that match empirical data and, then, to infer policy priorities via the associated endogenous (simulated) allocation of resources. The model of choice has been developed by Castañeda, Chávez-Juárez, & Guerrero, 2018) (henceforth referred to as the CCG model). Although describing highly specific details about the CCG model is beyond the scope of this article, we provide a sketch of its socioeconomic mechanisms, and concentrate on developing an empirical strategy that exploits its simulated policy priorities.

We frame our application in the context of the Organization for Economic Co-operation and Development (OECD). This is an appropriate setting in which to study policy coherence since, arguably, OECD membership is conditional on certain degree of alignment with a set of principles established by the existing members. Therefore, we argue that countries that joined the OECD after 1990 had incentives to follow the lead of the early OECD members (i.e. to emulate them to some degree). The core of our analysis concentrates on Mexico, an upper-middle-income country (UMIC) that, in the official government discourse, has advocated coherent policies.[3] Such discourse is not unique to Mexico's last administration, but has been a defining feature of every government over

---

[3] "A National Council for the 2030 Agenda for Sustainable Development, chaired by the president, was established in 2017. Its main purpose is to coordinate the actions for the design, execution and evaluation of [...] policies [...] for the compliance with the 2030 Agenda" (OECD, 2018, p. 135). The government website www.gob.mx/agenda2030 contains a repository of documents and information regarding the SDGs in Mexico.

the last 30 years. Despite this apparent enthusiasm for 'catching up' with the original OECD members, our findings suggest that Mexico's policy priorities have not been coherent. Even worse, we find that its priorities have been quite the opposite of what they would be if the government were serious about reaching these goals. For scholars familiar with Mexican development these results are not surprising. In fact, they serve to validate the proposed index. We reinforce such validation by analysing two additional country cases: South Korea and Estonia. Overall, we find that our results are consistent with development narratives and qualitative evidence from experts on these nations.

The rest of the article is structured in the following way. In section 2, we discuss the concept of coherence with regard to policy priorities and present a literature review on alternative approaches. Section 3 presents the data and methods to estimate policy priorities through the CCG framework. In section 4, we build the coherence index and present our main empirical findings. Finally, in section 5 we discuss the limitations and potentials of our approach, and provide some conclusions.

2 ON THE COHERENCE OF POLICY PRIORITIES

The idea of policy coherence for development has lingered in academic research and policy reports for some decades. While development economists have not been particularly interested in this concept, a large number of practitioners and academics in the fields of development studies, policy sciences and public administration have extensively discussed the idea. Broadly speaking, coherence has been qualitatively studied in the literature on Policy Coherence for Development (PCD) (Forster & Stokke, 2013). Originally, PCD was conceived as a principle for the international aid system. The main idea was that donor countries should also consider the impact that all policies established for their own benefit have on the development of poor countries, and not only the effects on development aid policies in those nations (Sianes, 2017; Barry, King, & Matthews, 2010). PCD, however, has evolved to the point of becoming an evaluation standard for the planning and implementation of policies in countries trying to achieve a set of development goals, irrespective of whether their economies are advanced, emerging or poor. Arguably, one of the main drivers in the widespread use of this standard has been the OECD through its several reports on policy coherence. For instance, the opinion expressed by the Secretary-General Angel Gurría in the foreword of its 2018 report (OECD, p. 3) provides an up-to-date notion of this concept:

> *[PCD] calls for breaking out of policy silos and increasing capacities to identify, understand and manage interactions and interconnections among SDGs. It entails harnessing*

synergies, managing trade-offs and policy conflicts, and addressing the potential transboundary and intergenerational policy effects of domestic and international action.

This definition is an important step in the right direction for the PCD framework because it makes explicit the importance of policy–policy interdependencies and policy–goal interactions. With such high-level official recognition, several development analysts have reframed PCD as a systemic problem. Consequently, it has become prevalent among international organizations and academics to identify potential reinforcing and conflicting effects between development indicators. For example, when the OECD and other multilateral agencies (e.g. the United Nations Development Programme (UNDP) and the World Bank) extended their agenda from PCD to policy coherence for sustainable development in 2014, they aimed at integrating economic, social and environmental dimensions of development across all levels of domestic and international policy-making through their complex interdependencies. Unfortunately, the demand for identifying interactions between policies has exposed severe technical limitations in qualitative approaches; in particular, their heavy dependence on expert knowledge in highly specific fields, which precludes scalability and introduces conflicting biases (see Ospina-Forero, Castañeda, & Guerrero (2019) for a review). Therefore, developing systematic, quantitative and scalable frameworks has become an essential endeavour.

Before going into the proposed method, it is important to be aware that the development literature defines policy coherence at different levels and stages (see, for instance, Curran, Dougill, Pardoe, & Vincent (2018); Carbone (2008)). First, horizontal coherence alludes to the interactions between policy issues and how these make it possible to attain different goals simultaneously. Vertical coherence, in contrast, is used to describe the connections between policies at different tiers of government (e.g. regional and national, national and supra-national). Second, policy coherence can be analysed at two stages: design and implementation. The former relates to the formulation of policy priorities by analysts and policy-makers. The latter involves the coordination of different government actors responsible for the operational side of policies. In both classifications, assessing policy coherence requires a partnership between scientists of different disciplines and technocrats dealing with issues of public administration.

In this article, we develop an approach that emphasizes policy coherence at the horizontal level. This proposal focuses on the design stage and is based on modern scientific data-analysis tools. In this sense, our focus is more narrowly defined than PCD, which is an overarching term for different discussions on policy coherence. Hence, in order to prevent any confusion with the OECD definitions, we avoid using the term PCD altogether. Instead, we use the term *coherence of policy*

priorities or, more generically, *policy coherence*. Delimiting the scope of our study allows us to construct a more comprehensive definition of coherence; one that goes beyond the identification of positive and negative spill-overs, and helps in considering the specific constraints and inefficiencies that nations face in their processes of development.

2.1 Challenges in measuring policy coherence and related literature

In this section, we discuss some of the main challenges that need to be met in order to quantify policy coherence. We elaborate on four problems that, in our opinion, are inadequately dealt with in existing approaches. In addition, we believe that each of these challenges can be addressed through the CCG framework.

The first problem—implementation inefficiencies—relates to the limitation of directly observing coarse-grained government-expenditure data, but not the policy priorities behind official statistics and discourse. The second—spill-over effects—refers to the misleading practice of inferring coherence from the interdependencies of socioeconomic indicators. The third problem—context specificity—relates to the loss of a country's contextual information when pooling cross-national data for statistical analysis. Finally, the fourth challenge—implicit benchmark— alludes to the need for counterfactual analyses in order to generate country-specific reference points. Next, we discuss each challenge in detail. At the same time, we review some of the existing methods, highlighting their virtues and pitfalls.

2.1.1 Implementation inefficiencies

A recent study on public expenditure by the Inter-American Development Bank (IDB) reveals a major problem of misuse of resources in Latin America. Such wastage is the consequence of a lack of bureaucratic professionalism, negligence, corruption, or a combination of all three (Izquierdo, Pessino, & Vuletin, 2018). For instance, the study estimates that, on average, inefficiencies in just three policy areas (procurement, civil services and targeted transfers) account for 4.4% of the region's gross domestic product (GDP), and about 16% of the total government expenditure (p. 63). To put this into perspective, a similar expenditure with respect to GDP is, on average, allocated to health (4.1%) and education (4.8%). Besides these technical inefficiencies, there are important allocative inefficiencies arising from a poor distribution of resources across generations, government levels and policy issues.

A comprehensive measure of policy coherence should consider both technical and allocative inefficiencies (corruption being an important component of the former). Thus, in our view, analysing policy coherence should take political economy considerations into account, since it is otherwise implicitly assumed that expenditure data reveal the government's intentions on how development

targets will be reached. This could yield misleading policy prescriptions because the political economy plays an important role in determining the resources diverted for personal gain or how they are wasted in misuse. Consequently, governments adapt their budgets to the political economy, obfuscating the connection between their true priorities and public expenditure data. Furthermore, these dynamics preclude the estimation of policy priorities from single-period data.[4]

2.1.2 Spill-over effects

Multilateral organizations now acknowledge that development goals are part of an "indivisible whole" and so advocate for policy coherence in the planning process. In order to perform this task, early studies with a "systemic" focus associate coherence with the promotion of policies whose indicators show synergistic effects (positive spill-overs). They also discourage investing in issues that exhibit trade-offs (negative spill-overs) and obstruct the attainment of the desired targets. This is the case of numerous studies based on a network of interdependencies among development indicators (Le Blanc, 2015; Pradhan, Costa, Rybski, Lucht, & Kropp, 2017; Weitz, Carlsen, Nilsson, & Skånberg, 2018; Allen, Metternicht, & Wiedmann, 2018; Zhou & Moinuddin, 2017).

Incoming positive spill-overs not only imply reinforcing effects but also create side benefits that transform the incentive structure of the functionaries in charge of implementing the policies. Castañeda et al. (2018) show that the functionaries' contributions to their corresponding policies decline when they receive substantial spill-overs from other policies. Consequently, the potential benefits from investing in policy issues that are highly central in a network may be offset (or even reversed) by the negative incentives at the receiving end of the spill-overs. For this reason, it is indispensable to develop methods that allow balancing reinforcing effects and distorting incentives.

2.1.3 Context specificity

It is well known that context matters in the success of a particular policy package. Conventionally, qualitative studies deal well with context since their methods allow focusing on highly specific settings. In quantitative studies, however, data limitations often impede country-specific inference. This is the case, for example, of network-based studies. Here, cross-national data are pooled in order to estimate spill-overs between different policies (Ceriani & Gigliarano, 2016; Cinicioglu, Ulusoy, Ekici, Ülengin, & Ülengin, 2017; Czyżewska & Mroczek, 2014; El-Maghrabi, Gable, Rodarte, & Verbeek, 2018).

---

[4] This situation worsens if the data cannot be properly disaggregated into fine-grained types of expense (transformative or committed), policy issues (topics) and government sources (federal, state and municipal).

Measures of policy coherence aiming to guide real-world policies should consider country-specific spill-over networks. Recent developments in network science make it possible to estimate the complex structure of synergies and trade-offs from highly restrictive datasets (see Ospina-Forero et al., 2019 for a review). This does not mean that the network, by itself, is enough to assess policy coherence, since the data from which a network is typically built (development indicators) do not account for implementation inefficiencies.

2.1.4 Implicit benchmark

By context specificity we refer not only to a specific network of interdependencies or to a unique configuration of development indicators, but also to a country's aspirations or development goals. In quantitative terms, these aspirations are represented through specific values that a government wants to achieve for each of the relevant development indicators. Depending on the nature of these goals, the policy priorities may vary, and hence the level of coherence. For instance, the allocation of resources may change substantially if governments want to emphasize addressing environmental issues, security concerns, inclusiveness problems, or even achieving a Scandinavian development model. To the best of our knowledge, no quantitative study takes this into consideration. This is important because, while a government's development goals may be influenced by multilateral organizations, they are always adapted to meet a country's specific idiosyncrasies.

2.2 A new definition

The non-observability of policy priorities, the existence of diverse causal channels, the presence of spill-overs, the importance of context specificity, and the need to establish development goals make assessing policy coherence quite challenging. As a first step we propose an alternative definition of policy coherence that can be operationalized in a quantitative fashion. The definition builds on the idea that policy coherence can be measured only in the presence of counterfactuals. Koch (2018) has previously introduced the idea of using baseline counterfactuals to assess policy coherence. This approach is mostly qualitative, so it relies on substantial expert knowledge to build counterfactual result chains. In contrast, our approach relies on computational simulations to generate such counterfactuals. Before going into further details on how simulations can be used for this purpose, we define coherence of policy priorities as follows:

> *The policy priorities of country X are coherent with its goals T if the allocation of resources*
> *P destined to transformative policies is similar to the allocation Q that X would 'discover'*
> *by trying to reach T.*

The definition is agnostic about the theories of change and policy-making that underlie policy priorities. Therefore, given a computational model that simulates the process of policy prioritization,

this definition can be used to assess policy coherence. In a subtle way, this definition addresses the four challenges raised in section 2.1. First, the definition requires a pre-specified set of targets for the indicators. These could be hypothetical values of development indicators, or the observed levels corresponding to a different nation *Y* that *X* would like to emulate (also called the "*Y* development mode"). Second, it requires estimating the policy priorities *P* from historical data. Third, *Q* is the counterfactual, which means that it needs to be estimated from a model where *X* would set *T* as its development goals and then try to reach them. Fourth, the term "discover" implies a learning process for the government. Thus, the model used to generate *P* and *Q* should consider country-specific factors such as inefficiencies and spill-overs.

To provide more clarity, suppose that the goals *T* are given, for a specific country, by the latest indicator values available in a historical data sample. Then, we can estimate *P* with this information. Because *P* has been inferred in a retrospective fashion, it provides an inference about the historical priorities of the government (for the sampling period). Now, if *T* is changed to be hypothetical rather than being based on historical data, the counterfactual policy priorities *Q* could be inferred. These priorities give us a counterfactual view of what the government would do in a hypothetical situation. The salient feature of this definition is that it binds factual and counterfactual priorities. It also highlights the need for using computational tools to generate *P* and *Q*, as will be explained below.

## 3 DATA AND METHODS

3.1 Data

We use a sample of the data constructed by see Castañeda et al. (2018), which comprises 79 development indicators at the country level. They come from three sources: The World Economic Forum's Global Competitiveness Report, the World Development Indicators, and the Worldwide Governance Indicators—the latter two produced by the World Bank. The coverage consists of annual observations for 117 countries over the 2006–2016 period. These indicators were normalized between 0 and 1 and have been readjusted so that better outcomes translate into positive changes. Appendix A provides a statistical summary of this dataset.

The sample used in this study consists of 34 OECD countries: 22 that joined before 1994 (we call them *early members*) and 12 that joined after 1994. With exception of Iceland and Luxemburg, all OECD early members are in this sample. The period is the same as the original dataset.

3.2 Computational model

In order to infer policy priorities, we use the CCG model (Castañeda et al., 2018). Broadly speaking, this model generates synthetic development-indicator data by simulating a government that allocates resources to different policy issues in order to reach a set of development goals. Figure 1 provides a sketch of the model mechanisms. On the left-hand side, we have the government agent (top circle), which tries to improve its indicators in order to reach a set of pre-established goals (bottom circle). Every period, the government allocates resources (solid black arrow on the top left) to different public servants (top-middle circle). These agents are in charge of transforming those resources into public policies (solid black arrow on the top right). However, inefficiencies in this transformation may arise because the agents can benefit from the misuse of resources (e.g. diversion of public funds, shirking, corrupt public tenders). Thus, the resources that finally contribute to policy implementation produce outcomes and spill-overs to other topics (top-right circle). These, in turn, generate changes in the development process (bottom-right circle). The government tries to monitor and enforce the efficiency of policy-making through governance instruments (specifically, through monitoring corruption and the rule of law). Thus, with time, public functionaries learn how much inefficiency can be "tolerated" (short dashed arrow). Finally, the government adapts its allocations according to the perceived efficiency of the functionaries (long dashed arrow) and how far the indicators are from the development goals (long dashed line). The technical details on the CCG model and the tests for its external and internal validation are provided in Castañeda et al. (2018). What is important for this application is that, given the initial and final values of development indicators, the model simulates their trajectories until convergence. In other words, the CCG model provides a theoretically informed empirical account of why the indicators behave in the observed way. Therefore, simulated variables such as the allocation of resource are informative about real-world policy priorities.

Figure 1: Conceptual sketch of the CCG model
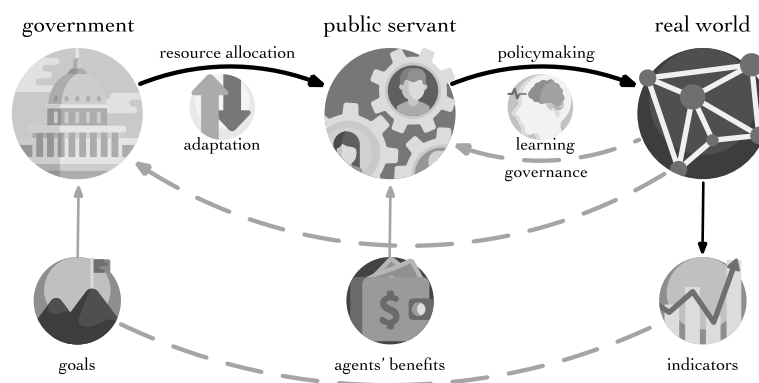
Source: Authors' calculations

The solid black arrows denote the chain of events between the government's allocation and the movement of the indicators. Grey vertical arrows indicate the incentives elicited by the development goals (for the government) and by the hidden inefficiencies (for the public servants). The dashed grey arrows and the dashed grey line show feedback effects that functionaries and the government take into account in their learning and adaptation processes.

3.3 Model usage

We can think of the CCG model as a software package that takes inputs and generates outputs. There are three inputs: (1) development-indicator initial values, (2) a network of spill-overs between indicators, and (3) development goals. Input 1 comes directly from the data. Input 2, the network, is obtained from Castañeda et al. (2018); more specifically, these authors estimate country-specific weighted networks for the indicators in our sample. Finally, input 3 consists of the aspirations of each country under study. In section 4, we explain how the latter type of input is determined.

Among the several outputs of the CCG model, we focus on simulated allocations. These are synthetic time series indicating the volume of resources that the policy issues received in each period. For a single simulation, each series is averaged across time. This gives us an average "allocation profile". Because this average profile may be sensitive to the agents' learning process, or to other random events of that particular realization, we produce 10,000 independent simulations. We then average their outputs to obtain an "expected allocation profile". This is the output that we use to construct the policy coherence index.

3.4 Context and spill-overs

Before proceeding to the analysis, we would like to take a detour to discuss the importance of the context specificity that should characterize policy design. In the CCG framework, context is given through its three inputs. First, the indicators tell us something about the stage of development of a particular country across multiple dimensions. Second, the network of spill-overs reveals how, for that specific country, those dimensions interconnect with each other. Third, the goals represent the nation's aspirations, which may be motivated by reasons that are highly specific to the context of that country. For example, a democratic government may be motivated by internal political agreements and the pressure of civil society, while a more autocratic one by emulating successful countries or by a broad international consensus (best practices).

The spill-over network is key to capture the context of each country. Castañeda et al. (2018) suggest that their structure varies drastically across countries at very different stages of development. Most importantly, these disparities are remarkable between nations which are relatively close in this regard (e.g. Chile and Mexico). In low-income countries, for example, improvements in poverty-

related issues seem to be associated with a diverse array of development pillars (e.g. education, health, governance). In UMICs, in contrast, improvements in public governance tend to condition changes in security, social inclusion and economic progress.

The context specificity of the network means that it should be estimated for individual countries, without pooling cross-national data. Ospina-Forero et al. (2019) review several methods that can be used for this purpose. For the analysis in this article, we employ the networks that Castañeda et al. (2018) have already estimated.

4 ANALYSIS AND RESULTS

The analysis of policy coherence in the CCG model requires defining the government's development goals (the third input). It is here when the counterfactual nature of our approach becomes evident. Let us recall that the period under study is 2006–2016. We then say that the retrospective government policy priorities are estimated via the CCG model using the following inputs: (1) the values of the indicators in 2006, (2) the spill-over network estimated for the sampling period, and (3) the values of the indicators in 2016. By establishing the goals as the final values of the empirical data, we are inferring the priorities a country established during the sampling period. In other words, the retrospective priorities are those inferred from historical data.

We could then establish some hypothetical goals while inputs 1 and 2 remain unchanged, creating a counterfactual simulation. For instance, if the hypothetical goals of country *X* were the indicators that another nation *Y* had in 2006, the inferred priorities would represent the counterfactual allocation that *X* would have established had its government tried to emulate *Y*. Since this type of counterfactual represents genuine intentions from *X* to emulate *Y*, we say that the associated policy priorities are the *consistent* allocation profile with respect to the *Y* development mode.[5]

Effectively, the retrospective and the consistent allocation profiles represent the terms *P* and *Q* (respectively) from our definition of policy coherence. Here, we have established a way to use the CCG model with real-world data in order to construct quantitative counterfactuals. Next, we present some results from applying these principles to the empirical dataset and, then, explain how to construct the coherence index.

---

[5] The logic of emulating a successful nation is the basic principle behind Akamatsu's flying geese (Akamatsu, 1962) and the idea of "development footprints".

4.1 Retrospective priorities

The left panel in Figure 2 shows Mexico's retrospective allocation profile, aggregated into 13 development pillars (based on those of the World Economic Forum). The aggregation serves for visualization purposes only, comprising the average priority given to the indicators for each development pillar. It is important to emphasize that the prioritization for the retrospective allocations is estimated at the indicator level.

In order to demonstrate that the data, by themselves, do not reveal policy priorities, we include three additional panels in Figure 2. These are direct "back-of-an -envelope" calculations such as those commonly done in benchmarking analysis. Clearly, the priorities that emerge from the CCG model do not resemble any of the other three panels. This suggests that the theoretically informed causal mechanisms specified in the CCG model produce non-trivial results. For instance, the top pillar in the allocation profile, technological readiness, is not even the second in any of the other data configurations. This not only illustrates the non-triviality of the estimations, but also of the importance of conducting policy evaluation through theoretically grounded models of the policy-making process.

Figure 2: Estimated and naïve allocation profiles for Mexico
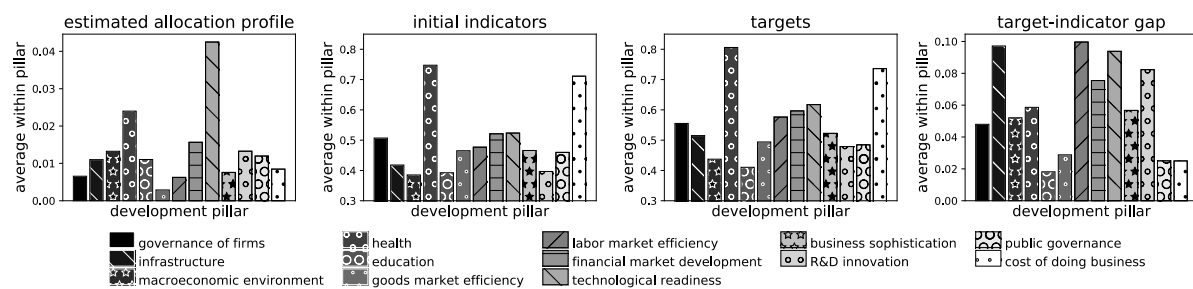
Source: Authors' calculations

All development-indicator data have been aggregated into 13 pillars by averaging the allocations of the indicators within each pillar. Panels from left to right: first, estimated allocation profile during the sampling period; second, initial development indicators; third, final indicators; fourth, differences between targets and initial indicators.

### 4.2 Counterfactual priorities

Let us return to the proposed framing of policy coherence with respect to emulating OECD countries. Here, we make counterfactual estimations for Mexico. The hypothetical goals in each counterfactual are the indicators (in 2006) of an OECD early member. In other words, the exercise simulates the priorities that Mexico would have established if, in 2006, its government had

genuinely wanted to catch up with an OECD early member. The OECD nation that better represents Mexico's aspirations is a question that lies beyond the scope of this article. Here, we are agnostic about preferred development modes, and perform counterfactuals for all OECD countries that joined before Mexico.[6]

Figure 3 compares the retrospective priorities (left panel) against the average counterfactual ones (right panel). For visualization purposes, we have aggregated all the counterfactual priorities (across all development modes) in the right panel. However, we use the disaggregated data to construct the coherence index and present the full results in Table 1 and further findings in Appendix C. Clearly, the priorities inferred from historical data differ from those that would be consistent with emulating exemplary OECD nations. For instance, the right panel suggests that Mexico would have to invest heavily in public governance in order to catch up with the OECD early members, while it currently prioritizes technological readiness. Note that the top-five priorities for Mexico in the consistent profile, at the pillar level, are public governance, education, R&D innovation, cost of doing business and health. We exploit the discrepancy between retrospective and consistent profiles to build a metric for the coherence of policy priorities.

Figure 3: Retrospective and consistent allocation profiles for Mexico
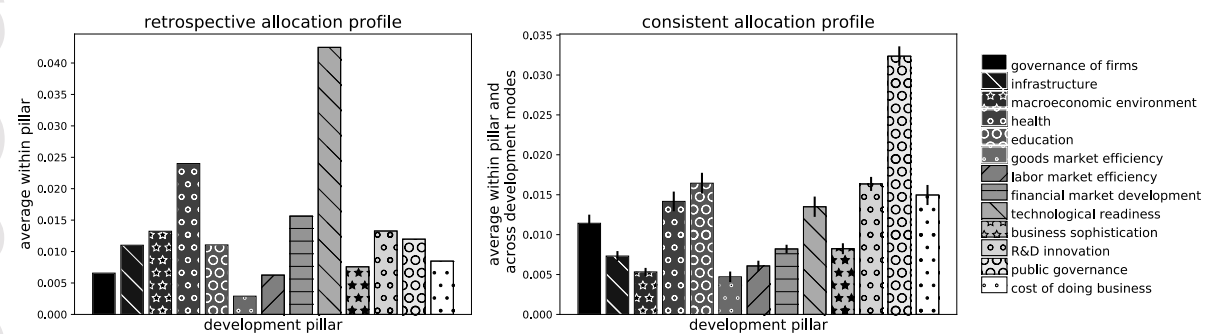
Source: Authors' calculations

All development-indicator data have been aggregated into 13 pillars. The vertical black lines at the top of each bar of the right panel denote the standard errors from the cross-national aggregation of consistent allocation profiles.

---

[6] The question of how a government determines its goals lies beyond the scope of this article and is quite a complex problem since it means turning the third input of the CCG model into an endogenous variable. This would require extensive knowledge and data of the social mechanisms through which nations shape their aspirations. To the best of our knowledge, there currently is no quantitative model that can explain how countries generate development goals.

4.3 A coherence index

In order to build the coherence index, we require some basic mathematical notation. Recall that P and Q correspond to the retrospective and the consistent (counterfactual) policy priorities of a given country.

Effectively, these priorities can be represented as vectors or numerical lists. We illustrate how these lists differ for the case of Mexico across three counterfactuals in a disaggregated fashion. Figure 4 plots the retrospective allocations in the horizontal axes, and the consistent ones in the vertical axes. The three panels illustrate the counterfactuals of Mexico emulating Turkey, Italy and the USA respectively (chosen for illustration purposes only). Let us concentrate on one of these examples, say the USA. Graphically speaking, full policy coherence would mean that all the dots in the right panel would lie on the dashed line, i.e. both the retrospective and the consistent priorities are identical. Deviations from the 45-degree line denote discrepancies between the allocation profiles and, consequently, a lower level of policy coherence. When a dot lies on the left side of the 45-degree line, it means that, in the retrospective allocation, this particular indicator received less priority than it would have received in the counterfactual. This is an allocative inefficiency of under-expenditure. The same logic applies for dots that lie to the right of the dashed line. These correspond to allocative inefficiencies of over-expenditure.

Figure 4: Mexico's coherence: retrospective versus consistent allocations for Mexico
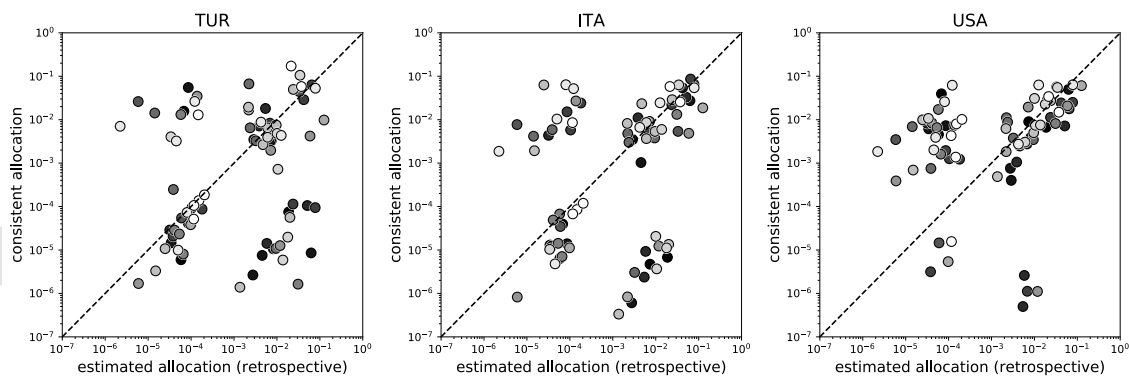
Source: Authors' calculations

Three examples of development modes that Mexico can adopt from the early OECD members. Full coherence implies that each dot representing a particular policy issue should be along the 45-degree line.

Measuring allocative inefficiencies (the distance between a dot and the 45-degree line) gets us closer to measuring policy coherence. Nevertheless, we also need a reference point for how incoherent a country can be with respect to its development goals. This additional counterfactual

can be obtained by inverting Q. That is, we can construct a new vector R where the most prioritized indicator is the least prioritized one in Q. The second most prioritized in R is the second least prioritized in Q and so on and so forth (we preserve the magnitude of the priorities). Thus, R is a mirror image of Q, and it gives us the reference point for full inconsistency with respect to a given development mode.

Now that we have P, Q and R, we need to measure their differences. A popular approach to quantifying the distance between two numerical lists X and Y is to add up the absolute value of the differences between their elements. For instance, $|X_i - Y_i|$ computes the difference between the element ith of X and Y and gets its absolute value. Therefore, the formula $d = \sum_{i=1}^{n}|X_i - Y_i|$ performs this operation for all the paired elements from X and Y, and returns the total sum as a measure of the distance between both vectors. For the remainder of the article, let us use this distance metric. We provide robustness tests for other metrics in Appendix C.

The coherence index consists of measuring the distances between P and the other counterfactual priorities Q and R. By arranging these measures as

$$h = \frac{d(P,R) - d(P,Q)}{d(P,R) + d(P,Q)}$$

obtaining an index that behaves like a correlation coefficient, ranging from -1 to 1. If the index is negative, it means that the country is incoherent with respect to the given targets (because the retrospective priorities are more similar to the inconsistent counterfactual). If it is zero, it means that coherence/incoherence is ambiguous. When h is positive, we say that the policy priorities are coherent (because the retrospective priorities are more similar to consistent counterfactual). Finally, if $h = 1$ or $h = -1$, we speak of full coherence or full incoherence. One would rarely expect a coherence index of 1, since all of a country's policy issues are still far from being fully captured by existing datasets. Therefore, the index should not be interpreted for its magnitude alone, but by its statistical significance and sign; and should be used to compare it against other counterfactuals or against other countries' indices.

In order to speak of statistical significance, we need to test whether or not $h = 0$. For this, it is necessary to obtain the distribution of h, which can tell us if $h = 0$ would be expected with certain level of confidence. Recall that the P and Q are obtained by performing $2 \times N$ independent simulations of the CCG model, and that h comes from their averages. Therefore, in order to obtain a second estimate of h, one could perform N new simulations for P and another N for Q. Then, this can be repeated multiple times in order to get a large sample of the coherence index. Such a method, however, could be computationally prohibitive. An alternative approach is to perform a

bootstrapping procedure, taking the sample of N retrospective profiles P and re-sampling them with replacement. We do the same for the sample of counterfactuals Q. For these new samples, we compute the expected allocation profiles, we construct R, and we compute h. We repeat this as many times as needed to construct the distribution of h. Here, we only need to run the CCG model 2 × N times, while the rest of the computations are relatively fast. Then, we can compute p-values in order to determine if the coherence index is significantly positive or negative. Our experimental tests suggest that this bootstrapping procedure produces similar distributions as those obtained through direct simulations.

4.4 Results for Mexico

We find that Mexico's policies are not coherent during the sampling period. Figure 5 shows the coherence indices estimated for the 22 possible counterfactuals from OECD early members. The index is negative in all cases. In addition, there is no association between the level of development of the country to emulate (here measured in income per capita) and Mexico's coherence. This suggests that, besides not being coherent, Mexico's government made no systematic effort to follow the most developed countries. In fact, within Mexico's different levels of incoherence, it seems that Italy's development mode yields the closest counterfactual priorities to Mexico's retrospective ones. Furthermore, other less-developed OECD countries such as Greece and Turkey are not associated with the lowest levels of incoherence, as is desirable for a nation that aspires to catch up with more developed countries.

Figure 5: Mexico and its coherence with OECD development modes
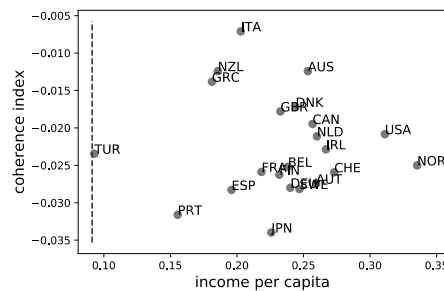
Source: Authors' calculations

Note: The dashed line denotes the Mexico's income per capita.

Figure 6 shows indicator-level allocative inefficiencies ($P_i - Q_i$) for the most (Italy) and the least (Japan) coherent counterfactuals. A negative difference denotes underspending while the opposite sign means over-expenditure. As shown in the two panels, Mexico has been underspending in most policy issues, yet it exercises disproportionate over-expenditure in the policy issue of *redundancy costs* (a component of the labour market efficiency pillar). Other two important topics where the

Mexican government overspends are *tuberculosis cases* and *general government debt*. Apparently, the burden of public debt, transaction costs in the labour market and expenses associated with certain diseases have become a hurdle for the implementation of more coherent policies. These results do not mean that expenditures should be cancelled, but rather the need to address the problems associated with these issues (assuming that Mexico's aspirations are well captured in the relevant development mode).

Although both development modes exhibit a similar pattern of allocative inefficiencies, we can pinpoint some differences. For instance, there are more cases of marginal over-expenditure in the Italian mode (e.g. in *R&D and innovation*). In the Japanese mode, underspending in *availability of the latest technology* and *extent of staff training* is more prevalent than in the Italian one. These differences show that attempting to reach Japanese targets requires a more technological focus and better human capital. They also illustrate that the distribution of inefficiencies across policy issues depends on the type of development mode that a country wants to pursue. In contrast, technical inefficiencies (e.g. corruption) depend on the strength of governance indicators.

In principle, some of the resources required to foster underspent policy issues could come from unnecessary allocations in other areas. Nevertheless, from Figure 6, it is clear that there are numerous underspent policy issues in both examples. For instance, investment in pillars like *public governance*, *R&D and innovation*, *health* and *costs of doing business* (principally in crime and violence) are especially large in the Italian and Japanese modes. Therefore, a major effort should be directed by the Mexican government to create the budgetary leeway for financing these costs and, as stated above, this starts by reducing the debt problem.

Figure 6: Mexico's allocative inefficiencies for two development modes
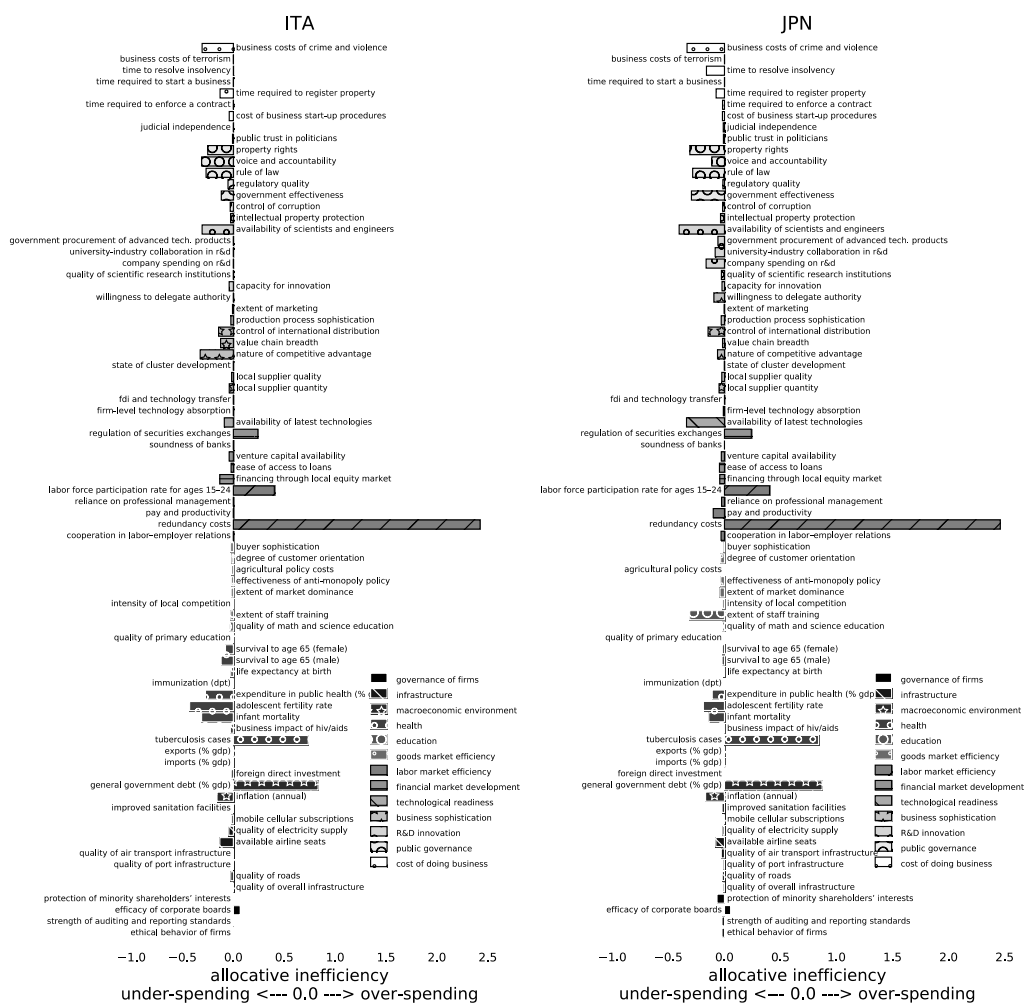
Source: Authors' calculations

The units in the horizontal axis have been re-scaled for presentational clarity. Each decoration denotes a development pillar. Positive bars indicate over-expenditure, while negative bars under-expenditure.

4.5 Validation

To scholars familiar with Mexican politics and development, the results obtained in the previous section are not surprising. In fact, they align well with a popular rhetoric on failed development strategies during the last three decades, providing some level of validity to the coherence index. In this section, we provide further validation cases by looking at countries whose development stories are better known, and that fit well within the context of joining the OECD and trying to emulate exemplary members. In contrast to the Mexican case, such stories are rather successful, demonstrating that the index is able to capture both coherence and incoherence in the real world.

The first country is South Korea, which joined the OECD in 1996. Knowing that it is a successful referent of development, and that its policy-makers have explicitly emulated numerous aspects of Japan's development mode (Lee & Yamazawa, 1996; Hsu, Naoi, & Zhang, 2014), validation should follow if: (1) South Korea has predominantly positive and significant coherence indices and (2) Japan is among the development modes with which South Korea is most coherent.

The left panel in Figure 7 confirms the validity of our method. First, note that all the indices are positive, and indeed most are statistically significant (see Table 1). Second, the Japanese development mode has a prominent position. This is consistent with well-known studies of the economic development of South Korea: "*Japan's development strategies have served as a model for Korean policy-makers*" (Lee & Lee, 2007, p. 13). Moreover, other highly coherent modes include Germany and Sweden, which speaks of the South Korea's visible transformation towards an outward-looking and technologically-oriented economy.

Figure 7: Two validation cases
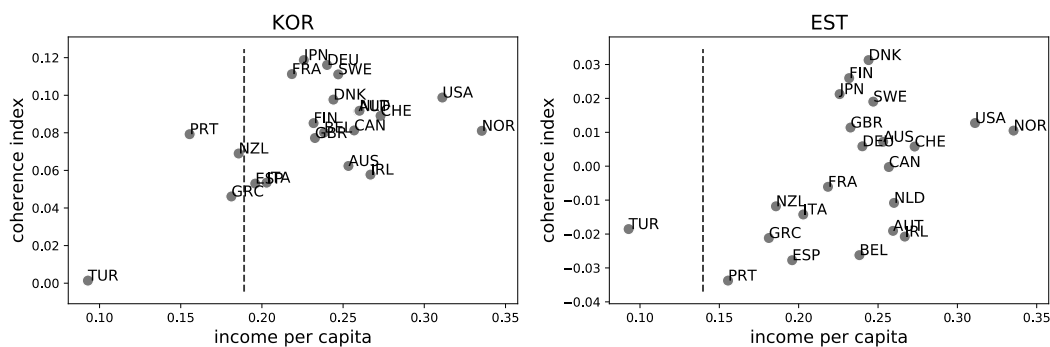
Source: Authors' calculations

Note: The dashed line denotes the income per capita of the country under study.

The second validation case is Estonia, a Baltic nation that is undergoing structural transformation. In contrast with Mexico and South Korea, Estonia became an OECD member in 2010, halfway through our sampling period. Here, our argument is that, prior to joining, Estonia was already undergoing a process of aligning its priorities to the requirements of the OECD. Hence, one would expect less coherence than in South Korea. However, the Estonian case is interesting because, after its separation from the former Soviet Union, it made serious attempts to adopt a Nordic development mode (Virkus & Harbo, 2002; Alestalo, Kuhnle, & Olsson, 2009). Thus, validity should be reflected in higher indices for the Nordic development modes. This is confirmed in the right panel of Figure 7, where Denmark, Finland and Sweden are three of the four modes with highest coherence.

Moreover, the other top country is Japan, which is consistent with the Estonia's impressive technological transition (Sinani & Meyer, 2004; Tiits, 2007).[7]

4.6 Results for other OECD members

To finalize our analysis, we present the results obtained for the other nine developing countries that became OECD members after 1994. This, of course, comes with the caveat that their reasons for joining the OECD are not as clear as for Mexico, South Korea and Estonia (for example, Turkey being a founding member in 1961 was clearly politically motivated). Table 1 shows all the estimates of h for each country-mode pair (highlighting the most coherent mode for each country). In contrast with a predominantly coherent South Korea, Slovakia is the most incoherent nation among the OECD late members (all but two counterfactuals have a negative value, and most are statistically significant). In addition, a country like the Czech Republic can be relatively coherent with one development mode (Sweden's) while, at the same time, be incoherent with another (Turkey's).

Poland and South Korea are the most coherent countries, with an average index of $h < 0.1$ across the 22 development modes. Their highest indices are in the order of 0.11. Furthermore, there are no indices in Table 1 that are positive and unambiguous at the 99% confidence level. Perhaps this apparent limitation of the coherence index can be overcome in the future as newer and bigger data can capture the full spectrum of policy issues that nations face. Nevertheless, this is an important first step towards the quantification of policy coherence. Appendix C presents similar results using different distance metrics.

Table 1: Coherence indices for augmented sample

Mode	CHL	CZE	EST	HUN	ISR	KOR	LTU	LVA	MEX	POL	SVK	SVN
AUS	0.1*	0.03	0.01	-0.01	-0.04	0.06	0.03	-0.03	-0.01	0.1*	-0.05**	0.05
AUT	0.07	0.07	-0.02	-0.02	-0.04	0.09**	0.01	-0.01	-0.03	0.1*	-0.04*	0.06
BEL	0.04	0.1*	-0.03	0.02	-0.03	0.08*	0.02	-0.02	-0.03	0.11*	-0.05**	0.06
CAN	0.09*	0.04	0.00	0.00	-0.04	0.08*	0.01	-0.01	-0.02	0.1*	-0.04	0.06
CHE	0.08*	0.09*	0.01	0.00	-0.02	0.09*	0.02	-0.01	-0.03	0.09*	-0.03	0.06
DEU	0.09*	0.1**	0.01	0.01	-0.03	0.12**	0.01	-0.01	-0.03	0.11*	-0.05**	0.06
DNK	0.07	0.1**	0.03	0.00	-0.02	0.1*	0.04	-0.02	-0.02	0.09*	-0.04	0.06
ESP	0.01	0.01	-0.03	-0.04	-0.04*	0.05	0.01	-0.03	-0.03	0.1*	-0.04*	0.02
FIN	0.08*	0.11**	0.03	0.02	-0.02	0.09*	0.04	-0.01	-0.03	0.1*	-0.04*	0.06
FRA	0.07	0.09*	-0.01	0.00	-0.02	0.11**	0.00	-0.03	-0.03	0.11*	-0.05**	0.05
GBR	0.11**	0.09*	0.01	-0.01	-0.03	0.08*	0.02	-0.03	-0.02	0.11**	-0.04*	0.06
GRC	-0.01	0.00	-0.02	0.07*	-0.01	0.05	-0.01	0.03	-0.01	0.07	0.00	0.02
IRL	0.07	0.06	-0.02	0.00	-0.04	0.06	0.01	-0.03	-0.02	0.1*	-0.05**	0.04
ITA	0.01	0.00	-0.01	0.02	-0.01	0.05	0.07	0.05	-0.01	0.06	0.03	0.03
JPN	0.06	0.08*	0.02	0.01	0.00	0.12**	0.05	0.01	-0.03*	0.09	-0.02	0.06
NLD	0.08	0.08*	-0.01	0.00	-0.03	0.09*	0.02	-0.01	-0.02	0.1**	-0.04*	0.06
NOR	0.1*	0.06	0.01	-0.01	-0.03	0.08*	0.02	-0.01	-0.03	0.09*	-0.04*	0.06
NZL	0.08	0.04	-0.01	0.00	-0.02	0.07	0.04	-0.02	-0.01	0.1*	-0.04	0.05
PRT	0.03	0.02	-0.03	0.03	-0.01	0.08*	0.00	-0.01	-0.03**	0.08*	-0.04	0.02
SWE	0.1*	0.09**	0.02	0.00	-0.02	0.11**	0.03	-0.02	-0.03	0.1*	-0.04**	0.06
TUR	0.01	-0.03*	-0.02	-0.03	-0.02	0.00	0.04	-0.03	-0.02	0.05	-0.01	0.01
USA	0.1*	0.1**	0.01	-0.02	-0.01	0.1**	0.04	-0.02	-0.02	0.09*	-0.04*	0.07

---

[7] It is also intriguing that there is a strong correlation between our coherence index for Estonia and the Inglehart-Welzel index for traditional secular/rational values in the 2010–2014 World Value Survey (Inglehart, Haerpfer, Moreno, Welzel, Kizilova, Diez-Medrano, Lagos, Norris, Ponarin, & Puranen, 2014). Perhaps, it is not so adventurous to hypothesize that a country's true development goals are largely defined on cultural grounds.

| Average | 0.07 | 0.06 | 0.00 | 0.00 | -0.02 | 0.08 | 0.02 | -0.01 | -0.02 | 0.09 | -0.03 | 0.00 |

Shaded areas are the development modes with maximum coherence for each country. Stars indicate statistical significance. *: p < 0.1, **: p < 0.05

Finally, using the entire sample, we would like to learn something about the average coherence of different development modes. That is, are there exemplary economies that developing countries coherently emulate in a systematic way? The left panel in Figure 8 shows that this is the case. Here, we have plotted the average coherence index of each development mode against its level of performance (measured through development indicators). The first thing to notice is that Denmark, Germany, Finland, Japan, Sweden and the USA are the development modes that are most coherently emulated—with the reservation that the overall level of the average index is still rather low. Nevertheless, we can also observe a clear positive association between how coherently a development mode is emulated and its economic performance. This suggests that our sample of developing countries tend to establish goals that resemble the indicators of the most advanced nations. The right panel confirms this result using income per capita as an alternative measure of performance.

Figure 8: Coherence across development modes
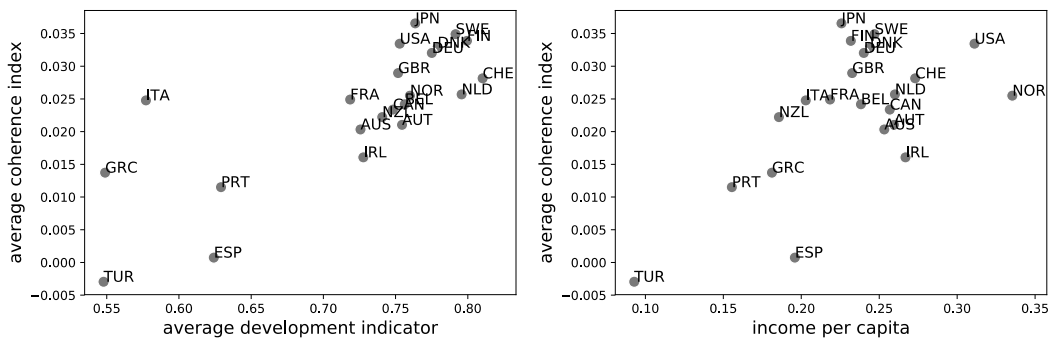
Source: Authors' calculations

The average is measured with the coherence index of all OECD latecomers when tracking a specific development mode.

5 CONCLUSIONS

In recent years, multilateral organizations such as UNDP, the OECD and the World Bank group, have actively promoted the SDG agenda, which states that countries must implement coherent policies in order to successfully reach a complex set of goals by 2030. As yet, however, there are no comprehensive quantitative methods to measure policy coherence.

This article introduces a new definition of policy coherence, based on the policy priorities governments establish in striving to achieve development goals. It also takes context specificity and political economy considerations into account. Using computational simulations, we construct counterfactual policy priorities, which we combine with historical ones in order to produce an index of policy coherence. The idea behind the index is that policy coherence should reflect a certain similarity between factual and counterfactual policy priorities (where the latter capture a government's stated goals).

When empirically estimating the index, we find that Mexico's development strategy has not been coherent with its rhetoric of emulating successful nations. This resonates well with well-known expert diagnostics of the country. We complement this empirical analysis with two country cases: South Korea and Estonia. The former has positive indices and a particularly high index for the Japan development mode, which is consistent with the development literature. Estonia, on the other hand, shows more coherence with the Nordic countries, which is also consistent with the literature. These case studies further validate the coherence index.

Overall, the proposed method provides a first approach to exploiting development-indicator data to develop a quantitative measurement of policy coherence. As more data become available, such estimates can be refined. Likewise, the CCG model could be further finetuned to specific institutional contexts. While this approach could help governments and organizations to produce quick and cheaper studies on policy coherence, this does not imply discarding other frameworks. Indeed, comprehensive studies on policy coherence should try to exploit the virtues of both quantitative and qualitative approaches in order to better capture the context of each country and, in turn, to improve the underlying data-generating model. This could prove extremely useful, for example, to assess whether governments receiving international aid have been coherent with how they are supposed to use it.

References

Akamatsu, K. (1962). A historical pattern of economic growth in developing countries. *The Developing Economies*, *1*, 3–25.

Alestalo, M., Kuhnle, S., & Olsson, S. (2009). *The Nordic model: Conditions, origins, outcomes, lessons*. Berlin: Hertie School of Governance. OCLC: 837213403.

Allen, C., Metternicht, G., & Wiedmann, T. (2018). Prioritising SDG targets: Assessing baselines, gaps and interlinkages. *Sustainability Science*, *14*(2), 421–438. https://doi.org/10.1007/s11625-018-0596-8

Barry, F., King, M., & Matthews, A. (2010). Policy Coherence for Development: Five challenges. *Irish Studies in International Affairs*, *21*, 207–223.

Carbone, M. (2008). Mission impossible: The European Union and Policy Coherence for Development. *Journal of European Integration*, *30*(3), 323–342. OCLC: 775276561.

Castañeda, G., Chávez-Juárez, F., & Guerrero, O. (2018). How do governments determine policy priorities? Studying development strategies through spillover networks. *Journal of Economic Behavior & Organization*, *154*, 335–361.

Ceriani, L., & Gigliarano, C. (2016). Multidimensional well-being: A Bayesian networks approach. Technical Report 399, ECINEQ, Society for the Study of Economic Inequality.

Cinicioglu, E., Ulusoy, G., Ekici, S., Ülengin, F., & Ülengin, B. (2017). Exploring the interaction between competitiveness of a country and innovation using Bayesian networks. *innovation and development*, *7*(2),1–36. https://doi.org/10.1080/2157930X.2017.1292617

Curran, P., Dougill, A., Pardoe, J., & Vincent, K. (2018). Policy coherence for sustainable development in Sub-Saharan Africa. Policy brief. Centre for Climate Change Economics and Policy. Leeds: University of Leeds.

Czyżewska, M., & Mroczek, T. (2014). Bayesian approach to the process of identification of the determinants of innovativeness. *Finansowy Kwartalnik Internetowy e-Finanse*, *10*(2),44–56.

El-Maghrabi, M., Gable, S., Rodarte, O., & Verbeek, J. (2018). Sustainable Development Goals diagnostics: An application of network theory and complexity measures to set country priorities. Policy Research Working Papers. Washington, DC: World Bank.

European Commission (2019). 2019 EU report on Policy Coherence for Development. Commission Staff Working Document SWD (2019) 20 final. Brussels: European Commission.

Forster, J., & Stokke, O. (2013). *Policy coherence in development co-operation*. Abingdon: Routledge.

Hsu, S., Naoi, M., & Zhang, W. (2014). Lessons in Sustainable Development from Japan and South Korea. New York: Palgrave Macmillan. OCLC: 881560466.

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, N., Norris, P., Ponarin, E., & Puranen, B. (2014). World Values Survey: All Rounds—Country-Pooled Datafile Version. Retrieved from http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp. Technical report. Madrid: JD Systems Institute.

Izquierdo, A., Pessino, C., & Vuletin, G. (Eds.) (2018). Better spending for better lives: How Latin America and the Caribbean can do more with less. Washington, DC: Inter-American Development Bank.

Koch, D.-J. (2018). Measuring long-term trends in Policy Coherence for Development. *Development Policy Review*, *36*(1), 87–110.

Le Blanc, D. (2015). Towards integration at last? The Sustainable Development Goals as a network of targets. *Sustainable Development*, *23*(3),176–187.

Lee, C., & Lee, K. (2007). Sustaining economic development in South Korea: Lessons from Japan. *The Pacific Review*, *5*(1),12–24.

Lee, C., & Yamazawa, I. (1996). *The economic development of Japan and Korea: A parallel with lessons*. New York: Praeger. OCLC: 966008569.

OECD (2015). *Better policies for development 2015: Policy coherence and green growth*. Paris: OECD. OCLC: 944185710.

OECD (2016). *Better policies for sustainable development 2016: A new framework for policy coherence*. Paris: OECD. OCLC: 957036944.

OECD (2017). *Policy coherence for sustainable development 2017: Eradicating poverty and promoting prosperity*. Paris: OECD. OCLC: 999369496.

OECD (2018). *Policy coherence for sustainable development 2018: Towards sustainable and resilient societies*. Paris: OECD. OCLC: 1040196886.

Ospina-Forero, L., Castañeda, G., & Guerrero, O. (2019). Estimating networks of Sustainable Development Goals. Working paper. https://dx.doi.org/10.2139/ssrn.3385362

Pradhan, P., Costa, L., Rybski, D., Lucht, W., & Kropp, J. (2017). A systematic study of Sustainable Development Goal (SDG) interactions. *Earth's Future*, *5*(11),1169–1179.

Sianes, A. (2017). Shedding light on Policy Coherence for Development: A conceptual framework. *Journal of International Development*, *29*(1),134–146. OCLC: 7770480682.

Sinani, E., & Meyer, K. (2004). Spillovers of technology transfer from FDI: The case of Estonia. *Journal of Comparative Economics*, *32*(3), 445–466.

Tiits, M. (2007). Technology-intensive FDI and economic development in a small country – The case of Estonia. *Trames Journal of the Humanities and Social Sciences*, *11*(3), 324–342.

Virkus, S., & Harbo, O. (2002). The internationalisation of Baltic library and information science education with emphasis on the cooperation with Nordic Partners. *Education for Information*, *20*(3-4), 217–235.

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

Zhou, X., & Moinuddin, M. (2017). Sustainable Development Goals interlinkages and network analysis: A practical tool for SDG integration and policy coherence. Hayama: Institute for Global Environmental Strategies. OCLC: 1015301713.

Appendix

A Data statistics

In order to provide initial summary statistics, we aggregate the indicators into 13 development pillars by computing the average value of all the indicators within a given pillar. In addition, we divide countries into three groups and a singleton. The first group consists of nations that joined the OECD prior to Mexico (before 1994). Arguably, some of these early members are exemplary nations on which Mexico has based its development goals for the last three decades. With exception of Iceland and Luxemburg, all OECD early members are in our dataset. The second group consists of countries with higher income per capita (IPC) than Mexico (it excludes those in group 1). These countries are useful for comparative purposes when describing the data. Group 3 is the singleton of Mexico. Finally, group 4 contains all countries with a lower IPC than Mexico.

Figure A1 displays the average level of development indicators of each group across the 13 development pillars. In order to convey information about group membership and development pillars in the same plot, we have assigned a double-pattern scheme to each bar. The bottom part of each bar is decorated according to the development pillar to which it belongs. This decoration also corresponds to the spatial grouping assigned to the pillars (each pillar has four bars). The upper smooth part of the bar denotes the group of countries for which the data is being averaged. Within each pillar group of four bars, the country groups have been sorted in the following order from left to right: group 1, group 2, Mexico, and group 4. Finally, the level of the bar indicates the average level of the relevant indicators.

The first feature to highlight from Figure 1 is that the level of the bars varies significantly across pillars and across groups. Note that, in general, the more advanced a country is, the higher its average indicators. For the Mexican case (light grey bars), the largest differences with respect to the OECD early members (black bars) correspond to the *education* and *public governance pillars*, while the smallest are in the *macroeconomic environment*, *cost of doing business* and *health pillars*.

Figure A1: Development pillars and income groups

Source: Authors' calculations
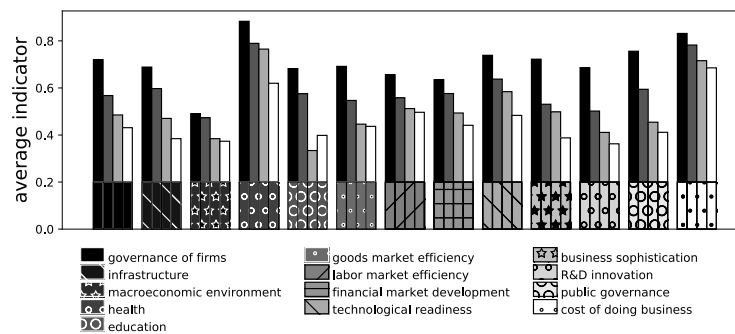
Average level of development indicators by pillar and group. The base decoration of each bar corresponds to a development pillar. Each bar within a pillar corresponds to a country group. That is, the black bars correspond to group 1 (OECD early members), the dark grey bars to group 2 (higher IPC than Mexico), the light grey bars to Mexico, and the white bars to group 4 (lower IPC than Mexico).

B Non-triviality of the coherence index

Figure 2 of the main text shows that the policy priorities inferred through the CCG model are not trivial because they do not correspond to back-of-the-envelope calculations. Nevertheless, the reader may wonder if this is also the case for the coherence index. Here, we show that the coherence index is not trivial. That is, that the results obtained in Figure 7 cannot be produced by simply plotting development indicators or their related back-of-the-envelope manipulations.

The top panels in Figure B1 try to replicate the results from Figure 7 from the main text, but using the Pearson correlation coefficient between a country's indicators and those from its development modes. Clearly both figures are different; for example, the German indicators are considerably less similar (DEU) to the South Korean ones than the Japanese indicators (JPN) and, yet, the European country is the second most coherent mode for South Korea. On the other hand, Ireland is the country with most similar indicators to Estonia, but its corresponding index falls into the domain of incoherence.

For the South Korean case, Japan is both the most similar and the most coherent. An interesting exercise would be to perform this test but for an earlier period in which South Korea was 'taking off'. Unfortunately, there are not too many indicators pre-dating the twenty-first century, so this is a task that we will leave for another paper.

The fact that the Japanese mode is both the most coherent and the most similar in indicators, begs the question of whether, among high levels of similarity between indicators, it is possible to find a positive association with coherence. The bottom panels in Figure A1 show that this is not the case. For instance, potential development modes for South Korea like Spain and Turkey with a high degree of similarity have a low coherence index with respect to the Asian country. Then, in the Estonian case, there are nine development modes with a similarity above 0.4, but whose coherence index is negative. Summarizing, the bottom panels do not show a positive relationship between the indicators-correlation and the coherence indices.

These results strengthen our advocacy for producing counterfactuals via computational models. They also point out the potential risks of naïve benchmarking. Thus, our work highlights the need to combine data-driven tools with theoretically-founded analyses.

Figure B1: The irrelevance of development-indicator comparisons for measuring policy coherence

Source: Authors' calculations
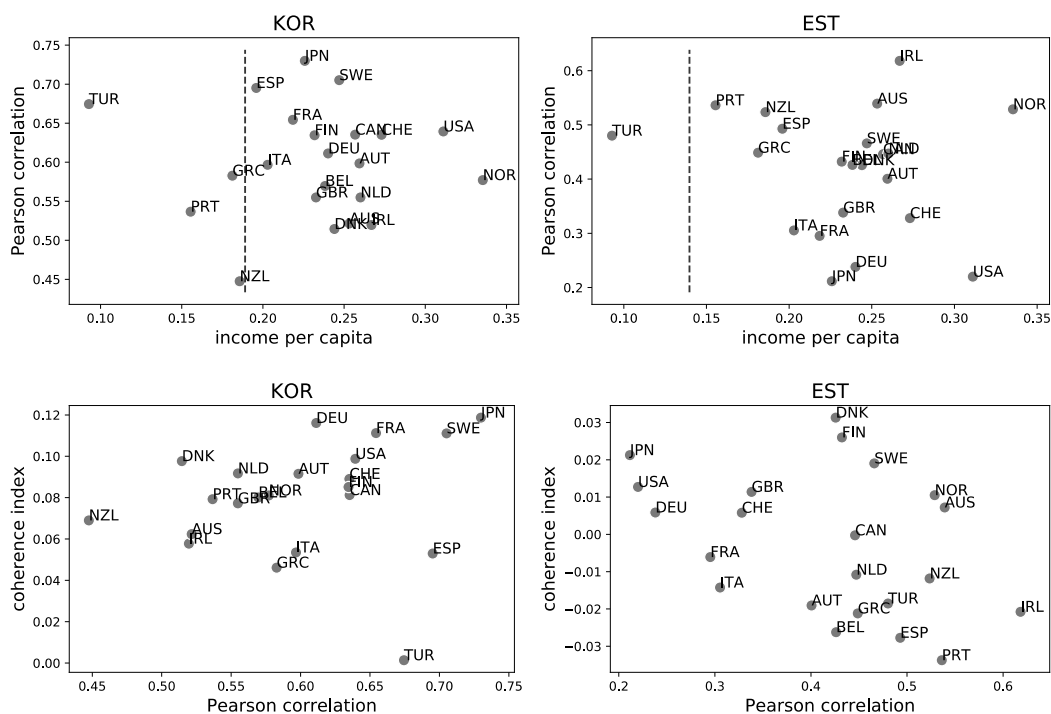
The dashed line denotes the income per capita of the developing country under analysis. We use the Pearson correlation as a similarity index between the country's development indicators and those of its potential development modes

C Robustness

Tables C1, C2 and C3 present estimates of the coherence index for the augmented sample under three alternative distance metrics: cosine, correlation and Euclidean distances. Overall, these tables demonstrate the robustness of our results. For example, Mexico remains ambiguous; Japan occupies a prominent position among South Korea's most-coherent development modes; and Estonia is most-coherent with the Nordic countries. These three alternatives (and popular) distance metrics are defined as follows.

Cosine distance is defined as

$$1 - \frac{X \cdot Y}{\|X\|_2 \|Y\|_2},$$

where $||\cdot||$ is the L2 norm $(\sum x^2)^{1/2}$. The correlation distance is

$$1 - \frac{(X - \bar{X}) \cdot (Y - \bar{Y})}{\|X - \bar{X}\|_2 \|Y - \bar{Y}\|_2},$$

and the Euclidean distance is the L2 norm for the difference between vectors X and Y:

$$[\sum_i (X_i - Y_i)^2]^{1/2},$$

Table C1: Coherence indices using cosine distance

Mode	CHL	CZE	EST	HUN	ISR	KOR	LTU	LVA	MEX	POL	SVK	SVN
AUS	0.17*	0.01	0	-0.04	-0.04	0.1	0.08	-0.04	-0.03	0.15*	-0.08**	0.05
AUT	0.15	0.09	-0.02	-0.06	-0.05	0.15*	0	-0.01	-0.05	0.15*	-0.06	0.07
BEL	0.08	0.16*	-0.04	0.02	-0.04	0.13	0.04	-0.03	-0.04	0.16*	-0.04	0.06
CAN	0.15	0.03	-0.01	-0.04	-0.05	0.13	0.03	-0.02	-0.04	0.15	-0.05	0.07
CHE	0.13	0.09	0	-0.03	-0.03	0.13	0.05	-0.02	-0.05	0.11	-0.05	0.08
DEU	0.15	0.15*	0	-0.01	-0.04	0.16*	0.01	-0.02	-0.05	0.17*	-0.07**	0.07
DNK	0.14	0.12	0.04	-0.05	-0.03	0.16*	0.08	-0.02	-0.03	0.13	-0.06	0.08
ESP	0.05	-0.01	-0.04	-0.08	-0.05*	0.1	0.02	-0.04	-0.05	0.15*	-0.04	0.01
FIN	0.15	0.14*	0.05	0	-0.03	0.12	0.06	-0.03	-0.04	0.14	-0.07*	0.07
FRA	0.1	0.16*	-0.01	0	-0.03	0.19*	0	-0.05	-0.04	0.16*	-0.05*	0.06
GBR	0.17*	0.11	0.01	-0.05	-0.03	0.11	0.03	-0.05	-0.03	0.17*	-0.06*	0.07
GRC	0	0	-0.02	0.14	-0.02	0.11	-0.05	0.05	-0.02	0.08	0.06	0.01
IRL	0.13	0.07	-0.03	-0.03	-0.03	0.11	0.01	-0.04	-0.04	0.15*	-0.06**	0.04
ITA	0.03	0	-0.03	0.05	-0.01	0.12	0.13	0.09	0	0.08	0.1	0.04
JPN	0.1	0.12	0.04	0.02	-0.02	0.18*	0.1	0.03	-0.05	0.13	-0.02	0.07
NLD	0.15	0.07	-0.02	-0.04	-0.03	0.15*	0.04	-0.02	-0.04	0.16*	-0.06	0.07
NOR	0.2*	0.05	0.01	-0.04	-0.04	0.15*	0.06	-0.02	-0.05	0.13	-0.06*	0.07
NZL	0.16	0.02	-0.02	-0.02	-0.02	0.13*	0.08	-0.03	-0.02	0.15*	-0.06	0.05
PRT	0.04	0.03	-0.05*	0.01	-0.01	0.17*	0	-0.01	-0.05**	0.11	-0.02	0.03
SWE	0.19*	0.12	0.03	-0.04	-0.03	0.19*	0.06	-0.03	-0.05	0.16*	-0.06*	0.08
TUR	0.02	-0.03*	-0.03	-0.04	-0.02	0.01	0.03	-0.07	-0.02	0.09	0.02	0.02
USA	0.16*	0.13	0.02	-0.05	-0.01	0.14*	0.09	-0.03	-0.03	0.16*	-0.04	0.08
Average	0.12	0.07	-0.01	-0.02	-0.03	0.13	0.04	-0.02	-0.04	0.14	-0.04	0

Shaded areas are the development modes with maximum coherence for each country. Stars indicate non-ambiguity significance. *: p < 0.1, **: p < 0.05

Table C2: Coherence indices using correlation distance

Mode	CHL	CZE	EST	HUN	ISR	KOR	LTU	LVA	MEX	POL	SVK	SVN
AUS	0.18*	0.01	0	-0.04	-0.04	0.1	0.08	-0.05	-0.03	0.15*	-0.08**	0.05
AUT	0.15	0.09	-0.02	-0.07	-0.05	0.15*	0	-0.01	-0.05	0.15*	-0.07	0.07
BEL	0.08	0.16*	-0.04	0.02	-0.04	0.13	0.04	-0.03	-0.04	0.16*	-0.04	0.06
CAN	0.15	0.03	-0.01	-0.04	-0.05	0.13	0.03	-0.02	-0.04	0.16	-0.05	0.07
CHE	0.13	0.09	0	-0.03	-0.03	0.13	0.05	-0.02	-0.05	0.12	-0.05	0.08
DEU	0.15	0.16*	0	-0.01	-0.04	0.16*	0.01	-0.02	-0.05	0.18*	-0.07**	0.07
DNK	0.14	0.12	0.05	-0.05	-0.03	0.16*	0.08	-0.02	-0.03	0.14	-0.07	0.08
ESP	0.05	-0.01	-0.04	-0.08	-0.05*	0.1	0.02	-0.04	-0.05	0.15*	-0.04	0.01
FIN	0.15	0.15*	0.05	0	-0.03	0.12	0.07	-0.03	-0.05	0.14	-0.07*	0.07
FRA	0.1	0.16*	-0.01	0	-0.03	0.19*	0	-0.05	-0.05	0.16*	-0.05*	0.06
GBR	0.17*	0.11	0.01	-0.05	-0.03	0.12	0.03	-0.05	-0.04	0.17*	-0.06*	0.07
GRC	0	0	-0.02	0.14	-0.02	0.11	-0.05	0.05	-0.02	0.08	0.06	0.01
IRL	0.13	0.07	-0.03	-0.03	-0.04	0.11	0.01	-0.05	-0.04	0.15*	-0.06**	0.04
ITA	0.03	0	-0.03	0.05	-0.01	0.12	0.13	0.1	0	0.08	0.1	0.04
JPN	0.1	0.12	0.04	0.02	-0.02	0.18*	0.1	0.03	-0.05	0.13	-0.02	0.07
NLD	0.15	0.08	-0.02	-0.04	-0.03	0.15*	0.04	-0.02	-0.04	0.16*	-0.06	0.07
NOR	0.2*	0.05	0.01	-0.04	-0.04	0.15*	0.06	-0.02	-0.05	0.13	-0.06*	0.07
NZL	0.17	0.02	-0.02	-0.02	-0.02	0.13*	0.09	-0.03	-0.02	0.16*	-0.06	0.05
PRT	0.04	0.03	-0.05*	0.01	-0.01	0.17*	0	-0.01	-0.05**	0.12	-0.02	0.03
SWE	0.2*	0.12	0.03	-0.04	-0.03	0.19*	0.06	-0.03	-0.05	0.17*	-0.07*	0.08
TUR	0.02	-0.03*	-0.03	-0.04	-0.02	0.01	0.03	-0.07	-0.02	0.09	0.02	0.02
USA	0.16*	0.13	0.02	-0.05	-0.01	0.14*	0.09	-0.04	-0.03	0.16*	-0.04	0.08
Average	0.12	0.07	-0.01	-0.02	-0.03	0.13	0.04	-0.02	-0.04	0.14	-0.04	0

Shaded areas are the development modes with maximum coherence for each country. Stars indicate non-ambiguity significance. *: p < 0.1, **: p < 0.05

Table C3: Coherence indices using Euclidean distance

Mode	CHL	CZE	EST	HUN	ISR	KOR	LTU	LVA	MEX	POL	SVK	SVN
AUS	0.09*	0	0	-0.02	-0.02	0.05	0.03	-0.02	-0.01	0.07*	-0.03**	0.02
AUT	0.07	0.04	-0.01	-0.03	-0.03	0.07*	0	0	-0.02	0.07*	-0.03	0.03
BEL	0.04	0.08*	-0.02	0.01	-0.02	0.07	0.02	-0.01	-0.01	0.08*	-0.02	0.03
CAN	0.07	0.01	-0.01	-0.02	-0.02	0.06	0.01	-0.01	-0.01	0.07	-0.02	0.03
CHE	0.06	0.04	0	-0.02	-0.01	0.06	0.02	-0.01	-0.02	0.05	-0.02	0.04
DEU	0.07	0.07*	0	0	-0.02	0.08*	0.01	-0.01	-0.02	0.08*	-0.03**	0.03
DNK	0.07	0.06	0.02	-0.02	-0.01	0.08*	0.04	-0.01	-0.01	0.06	-0.02	0.04
ESP	0.03	0	-0.02	-0.04	-0.03*	0.05	0.01	-0.02	-0.02	0.07*	-0.02	0.01
FIN	0.07	0.07*	0.02	0	-0.01	0.06	0.03	-0.01	-0.02	0.07	-0.03*	0.04
FRA	0.05	0.07*	-0.01	0	-0.02	0.1*	0	-0.02	-0.02	0.08*	-0.02*	0.03
GBR	0.08*	0.05	0	-0.02	-0.02	0.06	0.02	-0.02	-0.01	0.08*	-0.02*	0.03
GRC	0	0	-0.01	0.07	-0.01	0.05	-0.02	0.02	-0.01	0.04	0.03	0.01
IRL	0.06	0.03	-0.01	-0.01	-0.02	0.05	0	-0.02	-0.01	0.07*	-0.03**	0.02
ITA	0.01	0	-0.02	0.03	-0.01	0.05	0.06	0.04	0	0.04	0.05	0.02
JPN	0.05	0.06	0.02	0.01	-0.01	0.09*	0.05	0.01	-0.02	0.07	-0.01	0.03
NLD	0.07	0.04	-0.01	-0.02	-0.01	0.07*	0.02	-0.01	-0.01	0.08*	-0.02	0.03
NOR	0.1*	0.03	0.01	-0.02	-0.02	0.07*	0.03	-0.01	-0.02	0.06	-0.02*	0.03
NZL	0.08	0.01	-0.01	-0.01	-0.01	0.07*	0.04	-0.01	-0.01	0.07*	-0.02	0.02
PRT	0.02	0.01	-0.02*	0.01	-0.01	0.08*	0	0	-0.02**	0.06	-0.01	0.01
SWE	0.1*	0.06	0.01	-0.02	-0.01	0.09*	0.03	-0.01	-0.02	0.08*	-0.03*	0.04
TUR	0.01	-0.02*	-0.01	-0.02	-0.01	0.01	0.02	-0.04	-0.01	0.05	0.01	0.01
USA	0.08*	0.06	0.01	-0.02	0	0.07*	0.04	-0.01	-0.01	0.08*	-0.02	0.04
Average	0.06	0.04	0	-0.01	-0.01	0.07	0.02	-0.01	-0.01	0.07	-0.01	0

Shaded areas are the development modes with maximum coherence for each country. Stars indicate non-ambiguity significance. *: p < 0.1, **: p < 0.05