

MAKING POLICY EVALUATION WORK

THE CASE OF REGIONAL DEVELOPMENT POLICY

OECD SCIENCE, TECHNOLOGY
AND INNOVATION
POLICY PAPERS

April 2017 **No. 38**

FOREWORD

This paper was authored by Alexander C. Lembcke (OECD Directorate for Public Governance and Territorial Development) and Carlo Menon (OECD Directorate for Science, Technology and Innovation).

The authors wish to thank Paul Cheshire, Steve Gibbons, Nick Johnstone, Joaquim Oliveira Martins, Henry Overman, and Karen Maguire for useful comments on a previous draft.

The Committee on Industry, Innovation and Entrepreneurship (CIIE) declassified this paper on 5 November 2016.

Note to Delegations:

This document is also available on OLIS with the cote:
DSTI/IND(2016)4/REV1

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

© OECD 2017

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to rights@oecd.org.

ABSTRACT

The paper presents a critical discussion of ex-post impact evaluation of policies that affect regional economic development, with a particular emphasis on drawing useful implications for policy making. In particular, it discusses the importance of setting clear and measurable objectives in designing policies and the need for equally clear policy levers; it highlights the main advantages of “counterfactual” evaluation; it analyses the methodological specificities of the evaluation of programmes that have a regional or urban dimension; and it provides a survey of some of the most relevant examples in the empirical economics literature.

The ultimate goal is to “bridge” the perceived distance between policy discussions on the one side, and academic debates on the other. Some specific recommendations conclude the report.

TABLE OF CONTENTS

FOREWORD.....	2
ABSTRACT	3
MAKING POLICY EVALUATION WORK: THE CASE OF REGIONAL DEVELOPMENT POLICY ..	5
1. Introduction: Making evaluation useful for policy makers.....	5
2. “Regional” policies: Why place matters for policies.....	7
Lessons from economic theory.....	9
Agglomeration economies and transport costs.....	9
3. Setting the stage: Evaluation needs data, clear policy objectives and policy levers.....	11
What can be assessed: inputs, outputs and outcomes.....	11
Quantification requires data	12
Clear objectives and clear policy levers.....	13
4. Evaluating impact: What happens on the “road not taken”?	14
The chicken or the egg in policy impact: methodology matters!	14
Random assignment: the “experimental” strategy	15
Quasi-random assignment: getting close to the “gold standard”.....	16
5. Scale and spillovers: Challenges specific to the evaluation of “regional” policies	17
The choice of the right spatial unit and of the right data.....	18
The risk of a zero-sum game: Spatial substitution effects and sorting.....	19
Accounting for the first law of geography: Spatial autocorrelation and spatial spillovers	20
Not everything can be measured: Local unobservable factors	21
6. Good practice: Some recent benchmark evaluations	22
Transport infrastructure	22
Policies for lagging regions	23
Policies to attract investments	23
7. Conclusions and recommendations	24
NOTES	27
REFERENCES	30

Tables

Table 1. A stylised classification of regional development policies.....	8
--	---

Figures

Figure 1. Open Government data availability and accessibility	13
---	----

Boxes

Box 1. What is “regional”? Classifying subnational economic development policies	8
--	---

MAKING POLICY EVALUATION WORK: THE CASE OF REGIONAL DEVELOPMENT POLICY

1. Introduction: Making evaluation useful for policy makers

This paper presents a critical discussion of ex-post impact evaluation of development policies that have a spatial dimension, such as regional or urban development policies or, in short, “local” policies, with a focus on drawing practical and useful implications for policy makers. In particular, it discusses the importance of setting clear and measurable objectives in designing policies and the need for equally clear policy levers; it highlights the main advantages of “counterfactual” evaluation; it analyses the methodological specificities of the evaluation of programmes that have a local dimension, such as regional or urban development policies; and it provides a survey of some of the most relevant examples in the empirical economics literature. The ultimate goal is to “bridge” the perceived distance between policy discussions on the one side, and academic debates on the other.

This paper does not discuss the relative merits of specific policies *per se*. A quote from Rodrik (2008) motivates this choice: “Because most of the discussion to date has focused on the *whether* of industrial policy, the debate on industrial policy has reached diminishing returns and has become stale. [...] B]y focusing on the *how* of industrial policy we can move the debate forward.” The paper also takes a precise and narrow working definition of evaluation – i.e., the assessment of measurable outcomes of a programme in the short/medium term. This, however, is not necessarily the only component a policy maker (and a community) may be interested in, with qualitative and long-term considerations being also important and complementary, especially in the context of regional or urban development policies.

The speed and extent of technological advances in modern economies, the societal challenges linked to increasing inequality, and the growing importance of non-rival, knowledge-based production factors, are all phenomena that are likely to make targeted policy responses increasingly central in the political debate over the next years. The adjustment costs and frictions associated with these trends may simply be too fast and dramatic in some sectors, in some regions of a country or for specific groups (e.g., job displacement for low skilled workers due to automation) for public policy not to be expected to play a role.

At the same time, public resources are getting scarcer. This implies that there is a greater expectation of “value for money” in public policy, which requires a solid evidence base to inform policy makers on what works, and on what does not. There is therefore a pressing need for systematic improvement and more widespread application of public policy evaluation, including of policies aiming to support economic development.¹ With economic systems becoming more complex, the task of the evaluator is also becoming increasingly difficult, putting a premium on the use of the most robust methodologies.

Therefore, the methodological advances in impact evaluation techniques in economics over the last decade could not be timelier. This paper argues that there are huge benefits for policy makers from enabling and supporting rigorous ex-post impact evaluation of publicly-funded programmes that target specific sets of sectors, firms, regions or technologies. Rigorous evaluation by economists often shows mixed evidence and sometimes highlights the ineffectiveness of policies. But it is exactly for this reason that an “evaluation culture” is required, in order to identify successful and unsuccessful policies to support improvements and ensure that past failures are not repeated. More evaluation can lead to a deeper understanding of the mechanisms and the effects of public intervention in the market, possibly in conjunction with more policy experimentation.²

“Regional development policies” are taken here as policies that have a clear subnational dimension in their design, implementation or impact. They are used as the working example for three main reasons: first, they have played a prominent role among policy interventions in recent years; second, there is evidence suggesting that spatial inequalities are growing within OECD countries; third, development policies with a regional dimension have been systematically “underevaluated” in the past. To keep the discussion succinct, the focus is further restricted to policies with economic outcomes (such as employment creation, output or productivity growth). Despite these limitations, “regional economic development policies” remains a very broad category, but it does by no means capture all opportunities for improving policies through rigorous evaluation.³ Many conclusions of this paper can, however, inform and be applied to policies in general.

More than 70% of the European Union’s (EU) budget for Structural Funds has been devoted to supporting the economic convergence of lagging regions over the last decades and “regional development” is with nearly EUR 16 billion in 2014 the leading expenditure category of state aid by EU countries (European Commission, 2017). Regional policy is equally important at national level in many OECD member countries (OECD, 2016a). Despite an increasing focus of these policies on unlocking the growth potential of all regions, regional inequalities tend to be persistent.⁴ Rather than converging to an overall mean, regions seem to form “convergence clubs” that delineate the development in the best-performing regions from those lagging behind, each exhibiting their own level of wealth with significant differences across groups.

Some degree of inequality across regions will always exist. Some regions might benefit from natural resources that can be exploited, others from sights and scenery that attract tourists, others from good weather, or from an industrial structure that developed over decades (see e.g. Rappaport, 2004, or Cheshire and Magrini, 2006, for a discussion of the role played by climatic amenities). However, a pattern of within country divergence and international convergence is evident in the regional disparities within and across OECD countries. Over the past 20 years, labour productivity in large regions (TL2 regions⁵) across the OECD as a whole has slowly converged, but no such trend is evident within countries (Bartolini, Blöchliger and Stossberg, 2016; and OECD, 2016a). Similar patterns are found for the gross domestic product (GDP) per capita among large EU regions. Inequalities among regions⁶ in the EU27 and the EU15 have declined, albeit with some indicators showing little progress among EU15 regions after the mid-1990s. Conversely, within country inequalities have often increased (Magrini, 1999 and 2004; Monfort, 2008).

For cities and regions in the United States, the growing divide in economic well-being and related socio-economic indicators is forcefully described by Enrico Moretti as the “great divergence”. He maintains that this secular trend is “one of the most important developments in the United States over the past thirty years” (Moretti, 2012), and it not only concerns the United States but most industrialised countries, as it is rooted in the growing importance of innovation and the knowledge-based economy for productivity growth. This implies that governments in many OECD countries will be increasingly confronted with the need to address a growing spatial divide.

However, how to effectively address a “great divergence” is less clear, as impact evaluation of regional and urban development policies has attracted less attention in the economic literature than other areas of policy intervention, such as active labour market policies or educational policies.⁷ This is at least partially due to important methodological challenges that make evaluation in this field more complex. This report contributes to filling this gap by addressing explicitly the issue of evaluation of economic policies that have a local dimension, which encompasses a variety of policies that are expected to improve employment, productivity or growth.

The report is structured as follows. The next section outlines the scope of “regional” policies considered and provides a theoretical foundation of the importance of considering the local dimension for economic policies. The third section stresses some critical elements in the policy design phase. The fourth section presents a discussion of the salient methodological issues related to counterfactual impact evaluation techniques that focuses on being accessible to non-specialists. The fifth section highlights some methodological challenges specific to the evaluation of policies that have a local dimension. The sixth section describes some recent “good practice” examples from the empirical economic literature. The final section concludes and summarises the policy recommendations. Readers familiar with the more general challenges associated with policy evaluation may choose to omit sections three and four.

2. “Regional” policies: Why place matters for policies

A myriad of factors prevents people and firms from making use of opportunities that arise in an economy. Jobs outside a person’s commuting zone require costly relocation and can include indirect costs, such as the loss of local networks. The cost is often amplified by lock-in effects, e.g. the availability of cheap public housing or high transaction costs when selling property. Information about jobs is not universal, but often local and harder to obtain for more distant locations. Recent evidence from online job search in the United States shows that 90% of job applications are sent to firms within a 100 km radius and nearly all of them are within the applicants’ local labour market (Rathelot and Marinescu, 2014).

Whether policies should explicitly target particular places has been the subject of many (often heated) debates.⁸ On the one extreme, “place-based” policies are juxtaposed with “people-centred” policies, implying that only the latter consider the true target of an intervention. On the other extreme, policies without a place-based dimension are deemed “spatially blind”. These extremes miss the crucial point that many policies do have a “local” dimension (Box 1) and that explicitly taking the place-based dimension into consideration can help ensure that the policy’s impact on growth and welfare can be maximised (Garcilazo, Oliveira Martins and Tompson, 2010). Acknowledging the differential impact of policies across space allows the possibility to identify and leverage complementarities. For example, a country-wide voucher programme to support educational choice cannot succeed without ensuring that recipients of these vouchers actually do have a choice among local educational institutions.

Policies are often targeted specifically at places that are “lagging” according to some performance measures or places that face particular challenges (e.g. geographical location or industrial change). Examples for such policies can be found at small scales, e.g. the creation of enterprise zones or neighbourhood improvement initiatives, up to state or regional level, such as the EU Regional Development Fund. Successful policies should, however, consider the system of cities and regions in a country as a whole rather than target only “lagging” regions. Often the idea is that there is an economic imbalance that favours certain areas and by redistributing economic activity, jobs, or even governmental functions, a more equitable and less concentrated economy could be achieved. But even beyond the idea of overcoming excessive concentration, targeting struggling areas can help avoid vicious cycles, where a lack of opportunities leads to outmigration of the most skilled residents, which in turn further reduces opportunities. This requires a focus on local investment (over compensating subsidies) that aim to leverage each area’s unique growth potential and their competitive advantage. In practice, equity concerns are among the most prominent arguments for local interventions, together with a concern about spatial (skill) mismatch and the presence of “agglomeration economies” (Neumark and Simpson, 2015 provide an overview with a focus on industrial policy).

Targeting lagging regions may also imply that resources are not allocated to their most efficient use. Policies that artificially constrain mobility and the incentive for workers to seek the best possible job for them, can result in “misallocation” of talent, i.e. with the same effort the worker could contribute more in another job that might be in another location. This argument is even more relevant when unemployment

rates are very high and large discrepancies in unemployment arise across regions. Furthermore, policies that prioritise lagging regions might see part of their positive output counterbalanced by displacement effects, whereby the positive output in the targeted regions come at a cost of a symmetrically reduced output in neighbouring regions (see Enio and Overman, 2016, for a recent discussion). The discussion of agglomeration benefits in the following of the section is developing this issue further.

Box 1. What is “regional”? Classifying subnational economic development policies

A wide array of policies can be classified as “regional development policy”. It is a broad term that can be seen to encapsulate all efforts to support economic activity in regions and to reduce interregional disparities. It can also be seen to include policies that are designed to ensure equal treatment across space, but have a different influence on the economies in different regions or cities. It can therefore be useful to consider and make explicit the different dimensions in which they can be “regional”. Here, four dimensions are highlighted, but these are not exhaustive. Depending on the needs of the classification further dimensions could be incorporated, e.g. engagement of the civil society might be considered on its own; monitoring, operation or maintenance might be included. Similarly, this basic classification does also not distinguish between different levels of subnational actors. While not explicitly discussed here, extensions that account for different fields and levels of government are straightforward to add to this basic classification.

In the taxonomy of regional development policies proposed here, one or more of the following programme dimensions can be “regional”: the design; the implementation; the synergies, externalities, or market failures that the programme aims to address; and the expected effect of the programme. The first two relate to the level of government that designs or implements the policy. The other two dimensions capture regionally different impacts in policies due to market mechanisms and underlying economic fundamentals, which can be summarised as the synergies and externalities that arise; and the difference in outcomes across regions.

There are development policies which can be thought to be regional in all the four dimensions: e.g., a technology diffusion centre created by a subnational administration to foster knowledge diffusion within the metropolitan area. Policies of this kind are generally designed and implemented locally, they aim to activate synergies arising from localized interactions across economic agents, and the effect is also expected to be contained within the targeted metropolitan area. On the other extreme are policies that do not target local synergies but that create benefits throughout. An example is national infrastructure, such as broadband internet access or high-speed trains, which are typically designed and implemented nationally, and that do not specifically target local synergies or externalities. But even these policies are “regional” in the sense that their impact on spatial inequalities and on particular regions may be significantly different. Differences arise because the benefits (or costs) of a policy depend on existing location “fundamentals”, i.e. the tangible and intangible characteristics of a region or city. Improved access in an area with a thriving entrepreneurial community and a labour force with skills that facilitate the uptake of ICT in all parts of production will be able to create different benefits than a remote area where broadband could facilitate access to distance learning, remote consultation with specialist physicians or other consumption or service-delivery focused activities.

The classification remains of course stylised and in practice policies might have characteristics that combine different levels of government in e.g. the design (through consultation or joined development). The classification should nonetheless be useful to highlight some important distinctions and the breadth of areas that touch on regional development policies.

Table 1. A stylised classification of regional development policies

Regional aspects of policies	Design	Implementation	Synergies/ Externalities	Outcomes
Technology diffusion centres	✓	✓	✓	✓
Regional Development Agencies		✓	✓	✓
Cluster policies			✓	✓

Lessons from economic theory

Since the introduction of the “core-periphery model”, the seminal contribution of Paul Krugman (1991), there has been a wealth of theoretical contributions on the spatial organisation of economic activities, which are typically referred to as the “New Economic Geography”. This stream of literature developed somewhat independently from previous contributions in the field of “urban economics”, including Henderson’s (1974) model of optimum city size (Combes, Duranton, and Overman, 2015). Both strands are equally relevant for an exhaustive framing of the theoretical debate on the topic. In the following, the exposition will be limited to a few useful examples of some theoretical intuitions that are useful for policy design and evaluation. The interested reader can find a comprehensive and technically accessible treatment of related issues in Combes, Duranton, and Overman (2015).

Two main factors determine the geography of economic production: transport cost and local economies of scale or “agglomeration economies”, e.g. agglomeration benefits, productivity benefits from bringing a large number of workers and firms together. If transporting goods or providing services in different locations incurs costs and no scale economies or agglomeration effects exist, the spatial distribution of production would mirror the spatial distribution of consumers (“backyard capitalism”; cf. Ottaviano, 2009). However, the primacy of large urban agglomerations in both developed and emerging economies stands as uncontroversial evidence that economies of scale and agglomeration economies are anything but negligible.⁹ There is, however, also the fact that economic activity is not concentrated in a single “megacity”, which shows that agglomeration economies run two ways: agglomeration benefits increase productivity of workers in larger cities, but these workers also have to pay “agglomeration costs”. These costs can be pecuniary, e.g. rents and house prices are higher in larger cities, but also non-pecuniary, e.g. cities suffer from congestion, lack of greenspace or high levels of air pollution.¹⁰

Agglomeration economies and transport costs

A substantial body of theoretical and empirical work has aimed to establish the importance of “agglomeration benefits”, the positive impact of larger cities on their firms’ and residents’ productivity (OECD, 2015b; Puga, 2010).¹¹ The theoretical underpinnings for agglomeration economies go back to the 19th century (Marshall, 1890) and posit that the positive impact arises through three channels, “matching”, “sharing” and “learning” (Duranton and Puga, 2004; Carlino and Kerr, 2015). A larger pool of workers or firms means a “deeper” labour market, which results in a wider range of jobs for workers to apply to and firms to screen a larger pool of applicants. The result is an (on average) better *match* between worker skills and job requirements and therefore higher level of productivity. Secondly, a larger agglomeration of firms makes a place more attractive for suppliers and allows for greater specialisation in intermediate goods and especially services, as more firms *share* the same suppliers. Importantly, the *sharing* mechanism may also involve final goods – both in term of quality and variety – which may translates into additional utility for specific groups of consumers. Thirdly, proximity is essential for (in)formal knowledge flows. In a dynamic and knowledge intensive economy, mutual learning and adoption of “best practices” is greatly facilitated by proximity. This means that the skill and knowledge of other people in their vicinity positively affects the productivity of a worker, as they *learn* from each other and develop new ideas, goods and services.¹²

Urban economists are still debating whether the most significant economic transformations of the last decades, i.e. globalisation, diffusion of information and communication technologies (ICTs), de-industrialisation, and dematerialisation of production, are making agglomeration economies more or less important. Empirical observation shows that innovation and research and development (R&D) activities are more spatially concentrated than (manufacturing) employment (Audretsch and Feldman, 1996; Carlino et al., 2012). But Griffith, Lee, and Van Reenen (2011), using patent citation data, find that agglomeration effects are strong in research-intensive sectors during the 1975-99 period, however knowledge spillovers in other sectors tend to become progressively less localised. More recently, Li (2004) finds opposite results –

namely an increasing effect of borders and distance in containing knowledge spillovers (again measured with patent citations), associated with a trend of strengthening knowledge agglomeration over time since 1975. Similarly, the general discussion of the “Great Divergence” by Moretti (2012) is centred on the thesis that the importance of knowledge-intensive jobs for stimulating and sustaining growth is increasing and that larger agglomerations draw increasing numbers of skilled and highly-educated workers making the success of large cities self-reinforcing. More evidence on the issue would be useful for policy making, especially considering that the recent technological changes are having a strong impact on the nature of knowledge creation and spillovers.

Agglomeration economies are also at the forefront of criticism of place-based policies. As agglomeration economies create productivity benefits, resisting concentration of economic activity can result in foregone growth (e.g. McGill, 2011 and World Bank, 2009). Though a “space blind” approach is not always “space neutral”, as policies often favour areas with strong political lobbies, e.g. capital cities (McCann and Rodríguez-Pose, 2011), and policies that do not account for regional specificities run the risk of missing out on complementarities that are likely to be most evident and manageable at the local level (OECD, 2011). The consideration of sunk costs in transport and urban infrastructures may also create some path-dependency in the geography of policy intervention. The benefit of a place-based approach does however often depend on value judgements, e.g. on whether migration is considered as a social cost or not. If it is, then policies aimed at reducing economic disparities by targeting specific places, e.g. in order to create more economic opportunities in lagging regions, can be preferred. But the case of transport infrastructure also highlights that considering regions in isolation will not help in addressing all potential trade-offs and leverage opportunities, which require considering the region within a wider area and with its links to other regions and cities (OECD, 2016a).

If migration is not considered costly or “undesirable”, the dynamic of better-performing regions and the jobs they create can make targeting policies towards these regions and supporting mobility seem more promising than a focus on lagging regions. This however neglects that economically better performing regions face challenges in other dimensions, e.g. housing supply not keeping up with demand and rapid price increases, which means that a whole policy mix would be required that includes not only support to people, but also places, e.g. through urban planning (see Cheshire, Nathan and Overman, 2015, for a critical discussion). For instance, it is well-established that firms and workers in larger urban agglomerations are significantly more productive than firms and workers in rural areas or smaller cities (e.g., doubling the city size productivity increases by 2 to 5%, according to the economic literature; see De la Roca and Puga, 2017; Combes and Gobillon, 2015, for a survey on the topic and Ahrend et al., 2014 for recent evidence for five OECD countries). Conversely agglomeration costs seem to increase in roughly the same order of magnitude (see e.g. Combes, Duranton and Gobillon, 2016, for land prices in France or Ahrend and Lembcke, 2016, for local cost of living in Germany). For the policy maker, addressing these costs is often easier than aiming to raise agglomeration benefits directly. For example, by relaxing building restrictions in larger cities, cheaper housing would become available which increases the size of the agglomeration, raising average productivity and opening higher productivity jobs for new workers. But migration-focused policies have additional drawbacks that need to be carefully assessed. Local networks might create local agglomeration effects that could be destroyed through outmigration and continued outflows of residents might leave large parts of a country in decline, threatening political and social cohesion.

This is particularly relevant when considering transport links and captured in “core-periphery” models developed within the “New Economic Geography” framework. Since trains or motorways run both ways, increasing connectivity of the “periphery”, i.e. the area with less economic activity, with the “core” can have two different effects. On the one hand, the periphery gains better access to the goods, services and the agglomeration benefits of the larger core, which results in benefits to both firms and residents in the periphery. On the other hand, better connectivity also improves the access of firms in the core to the

residents in the periphery, which increases the competition for firms in the periphery. Depending on how much transport costs are reduced by the infrastructure, this could be welfare improving for the consumers but also reduce the local economic activity in the periphery. Since costs to provide goods and services from the core to the peripheral area are now lower, firms initially located in the periphery might relocate to the core region as the reduction in transport cost allows them to benefit from the core's agglomeration benefits directly while maintaining access to peripheral markets. The net effect depends on the interaction of a number of different factors, including the strength of the agglomeration economies, the size of the local and peripheral market, the transport costs of final goods and services and the cost and supply of local inputs, the local demand for final goods, etc. A detailed discussion of the relevant factors and mechanisms can be found in the contributions by Redding and Turner (2015); Teixeira (2006); and Puga (2002).

The trade-off between fully exploiting urban agglomeration benefits and reductions in transport costs, while containing spatial disparities is becoming increasingly relevant also for policies that are not inherently "local" in nature, given the importance of urban agglomerations for innovation and other knowledge-based economic activities that require frequent interactions and benefit from face-to-face contacts and knowledge spillovers.

3. Setting the stage: Evaluation needs data, clear policy objectives and policy levers

The first step in successfully evaluating policies is to identify what it is that should be assessed. This might seem obvious but in practice many policies are rather vague regarding their objectives and or even the levers that they aim to employ. In addition to clear objectives and levers, evaluation also required data that provides good measures for both the levers and the outcomes of the policy. Better data, makes the assessments more straightforward, the econometric estimates more precise, and the evidence of statistically significant (non-zero) effects more likely, to the benefit of both the academic and policy makers' community.

What can be assessed: inputs, outputs and outcomes

Often policies are assessed on inputs, such as how much money was spent, sometimes on intermediate outputs, i.e. what was the take-up, what was built, how many households, firms or regions received support. This paper argues that policies should instead be mainly assessed on the basis of its impact on individuals, firms, the economy, the environment, etc., i.e. the "outcomes" that are the ultimate policy objective. To ensure that evaluation of intended outcomes is feasible, good policies need to include a clear statement of these policy objectives. This also helps delineate intended from unintended outcomes when studying the policy's effects.

Prior to any evaluation, adequate indicators that capture the elements of a policy need to be chosen. It is useful to consider a typology of three types of indicators: input, output and outcome indicators. Input indicators measure the resources that are used in the implementation of a policy. They can be captured in the spending on a policy, e.g. the funds that are dedicated to a policy, such as the total spending from EU Cohesion Funds, or in terms of other resources dedicated to the policy, e.g. the number of employees, such as the number of full-time staff working in local business development centres. These indicators capture the "effort" dedicated to a specific policy.¹³ They are critical in establishing the costs that are associated with a policy, but are uninformative when it comes to the policy's benefits.

Output indicators summarise the results of a policy in terms of the means that a policy uses to achieve its intended effects, e.g. outputs of a policy that aims at creating industrial clusters are the extent of the physical area dedicated to the industrial cluster, its connection to local infrastructure, etc. The output of a road infrastructure project would be the road itself and its characteristics. Combined with input indicators, output indicators allow the assessment of "efficiency" of implementing policies, e.g. how many highway

kilometres were built with a given investment. Input and output indicators can therefore help identify whether policy incur relatively high costs or resources in certain regions, but even in combination the two indicators fall short of providing evidence on the benefits of a policy.

To assess the “effectiveness” of policies more than input and output indicators are required, specifically it requires indicators that capture the outcomes associated with a policy. Outcome indicators capture underlying objectives of a policy, i.e. what the policy aims to achieve. They can typically not be directly affected by the policy maker, but depend on the choices people and firms make. For example, a policy can create an industrial park, set taxation incentives and provide business support, but it cannot directly affect the number of firms within the park nor their success (the goals that underlie the investment). Similarly, infrastructure investment can be used create a new local road (output), but the uptake of the new road, i.e. the number of users is an outcome, as is the economic development along the road, the location of new firms, etc.

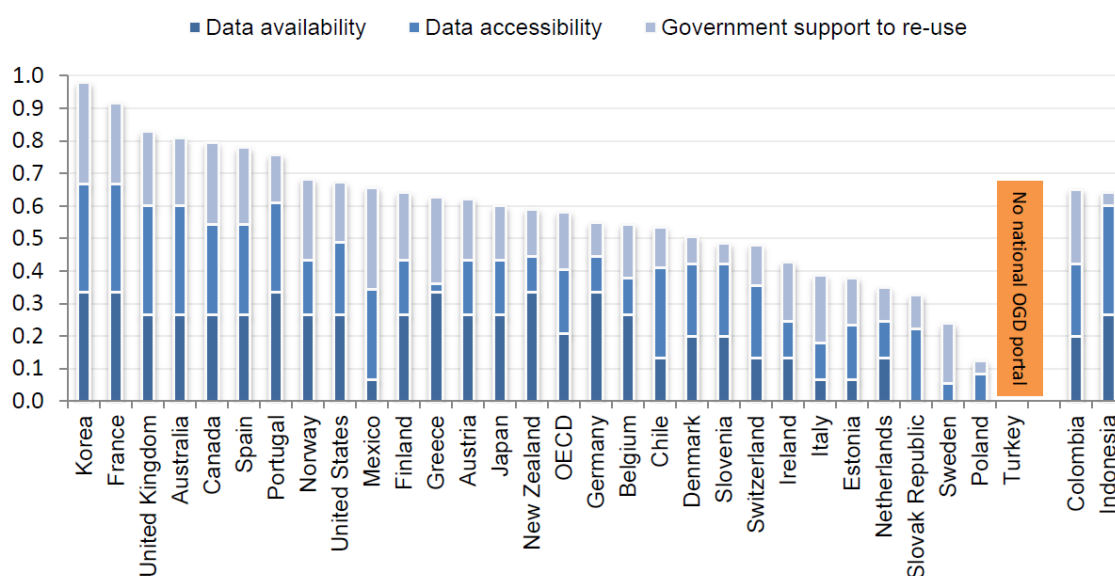
In addition to outcomes that are intended by a policy, a thorough assessment should consider indirect outcomes. The relationship between those who are directly affected by the policy and others may vary. For example, in some cases the effects of the policy on beneficiaries (i.e. new firms) may have no implications for others in the market (i.e. incumbents). Conversely in other cases a policy that benefits some people/firms/areas might simply result in a substitution effect, i.e., a crowding out due to the positive outcomes being counterbalanced by a specular negative outcome in neighbouring areas.

In the area of regional or urban policies these effects are endemic. Assessment of the implications for factor mobility, the interaction with local conditions and potential local externalities are crucial for the evaluation of policies. Often it is only possible to assess partial effects, but since the bigger picture might reveal substitution effects or complementarities, these assessments might have to be guided by theory.

Quantification requires data

The issue of data availability and data access has been relatively neglected until very recently. In recent years confidentiality rules and data protection legislation has been adapted in many OECD countries to allow better access to official large scale microdata sets. With the increased awareness of the potential to exploit “big data”, there has been some reconsideration of the importance of statistics and quantitative elaborations more generally. There is now an encouraging trend in many OECD countries in making more data freely available (open data), also in an effort of making the public sector more accountable (OECD, 2015e). This can be a crucial resource for good impact evaluation. However, it is important that some prerequisites are met. For instance, indicators should be based on data that is available in regular time intervals and they need to refer to an outcome that remains relevant for the foreseeable future. In most cases, regularity is no concern as the data usually becomes available at annual or higher frequencies. However, many surveys are one-off data collections that were not initially designed to be repeated on a regular basis. When considering such data sources for an indicator it should be taken into account whether it is possible to collect the data frequently enough to remain useful over the lifespan and potential impact period of a policy.

Often data exists within different parts of the administration, e.g. in tax records, school records or the unemployment registry, but legal barriers or a lack of incentives to utilise the data for assessments makes them practically inaccessible. Some OECD countries have made major steps in recent years to broaden access to confidential data, to link data from different sources, and to support the data re-use with appropriate documentation and information platforms. According to the 2014 OECD Survey on Open Government Data, Korea and France are the leading countries in this respect (Figure 1).

Figure 1. Open Government data availability and accessibility

Note: Data for the Czech Republic, Hungary, Iceland, Israel and Luxembourg are not available. Data for Indonesia are for 2015.

Source: Based on OECD 2014 Survey on Open Government Data and OECD (2015c), Government at a Glance 2015, OECD Publishing, Paris, http://dx.doi.org/10.1787/gov_glance-2015-en.

Clear objectives and clear policy levers

For a programme to be properly evaluated, a clear statement of the objectives should go together with an equally clear statement of the policy levers the programme is activating.

A peculiar challenge in creating local policies is that often clear and relatively uncontroversial policy objectives are confronted with a critical lack of hard evidence on the relative merits of the policy levers that should be activated to reach such objectives. Work on instrument choice is an open area of research in this domain. Cluster policies are a specific example for this. It is now established that the clustering of economic activities leads to productivity gains, although the sources and the underlying mechanisms of these gains are still debated in the economic literature (see Combes and Gobillon, 2015, for a survey). However, economic evidence on the specific policy tools that foster cluster formation, or those that enhance the productivity gains of existing clusters, is much scarcer and it might even be questioned whether policies can successfully replicate naturally occurring cluster benefits. As Duranton (2011) maintains, “due to the complex nature of the market failures at work it is unclear what cluster policies should do and how they should do it”. The main reason for this, according to the same author, is that clustering is an intermediate outcome, i.e., it is neither a “choice variable” which the policy maker can manipulate, nor the final desired outcome of the policy, which would rather be e.g. productivity growth, employment creation or innovation. Put differently, agglomeration economies and clusters tend to coincide in absence of policies because agglomeration economies push firms to cluster, but policies aiming to create clusters are not necessarily also triggering agglomeration economies. Efforts might therefore best be exerted in creating an attractive framework for firms and workers.

As such there is need for an understanding of the mechanisms which drive the observed relationship between the policy (the “choice variable”) and the cluster on the one hand, and between the cluster and the policy objective (e.g. innovation, employment, productivity) on the other hand. Therefore, when designing a policy at local level particular attention should be devoted to the fact that, as compared to other policy

areas, such as the labour market or education, our knowledge of how agents interact and the market works at local level is generally limited by the complex interactions among multiple agents. Furthermore, attention should be devoted to identifying the appropriate spatial scale at which the policies should be implemented (Nathan and Overman, 2013).

4. Evaluating impact: What happens on the “road not taken”?

*Two roads diverged in a yellow wood,
And sorry I could not travel both [...]*

Robert Frost (1916)¹⁴

Policies are not created in a vacuum but respond to observed or anticipated needs. This can confound the evaluation of a policy’s impact. If, for example, investment is targeted towards regions that have experienced a downturn in recent months, is the observed subsequent growth really caused by the investment or would the situation have ameliorated in the absence of the policy? Conversely, if no significant improvements can be observed, does this imply that the policy was unsuccessful or would the decline have further progressed without intervention?

Robert Frost’s poem signifies the conundrum of causal policy evaluation: if a person, firm or area is affected by a policy, it is impossible to observe what would have happened to the same person, firm or area if the policy was not in place. But it is exactly this difference that captures the *individual* causal effect of a policy (Rubin, 1974). To take a concrete example, the 2012 Olympics in London were promoted as part of an urban regeneration strategy for Stratford, a high unemployment neighbourhood in East London, and the whole lower Lea Valley area that stretches to the north and south of Stratford. To assess the impact that the investments and developments of the regeneration project had on employment in Stratford, it is necessary to observe the area both with the Olympics having taken place and without a 2012 Olympics.

As this is clearly impossible, policy research focuses on estimating *average* causal effects, rather than the individual causal effects on a given person, firm or area. The idea is that it is possible to classify otherwise similar people, firms and areas into groups: those that are affected by a policy and those that are not. Under certain conditions, the difference between the average outcomes in these two groups can then capture an average causal effect.¹⁵

*[...] And be one traveller, long I stood
And looked down one as far as I could [...]*

Robert Frost (1916)

The chicken or the egg in policy impact: methodology matters!

Different methodologies have been developed to identify a suitable “control” group for those people, firms or areas affected by a policy, i.e. a group that mimics that state in which those affected by a policy would have been in the policy’s absence. The unifying idea underlying the methodologies is that *ex ante*, i.e. before the policy is implemented, the control group is *on average* identical to the group that was affected by the policy. This means that it would not be possible to distinguish between the group that was affected and the unaffected group from one another before the policy came into place.

The challenge is that typically a myriad of factors confound the selection. Areas are selected because they exhibit particular characteristics, e.g. high unemployment rates, the location choice for an enterprise zone responds to a downturn in the established industries, etc. This means that the comparison of outcomes in an area before and after a policy was implemented will not capture the policy’s causal impact. But it also

means that it is often difficult to find suitable candidates to create a “control” group. If two regions suffer the same adverse shock to their industrial structure, why was one chosen as the location of an enterprise zone and not the other? Can it be ruled out that it is the same factor that drives any observed difference between the two areas?

Random assignment: the “experimental” strategy

Randomised controlled trials (RCTs) are often considered the “gold standard” of impact evaluation. Participants (e.g. people) are assigned to either the treatment or the control group, where the control group will be affected by the policy. Necessary for the validity of this strategy is that assignment to either group is random, which means unrelated to the policy or its expected impact on participants. With a large enough number of participants, random assignment will result in two groups with – on average – statistically indistinguishable differences in their characteristics. Whether randomisation was successful can (and should be) checked by “balancing tests”.

RCTs have become increasingly prevalent in the social sciences in developing countries (Duflo, 2006) and “[r]andomization has become an integral part of a development economist’s toolbox” (Duflo, Glennerster and Kremer, 2007; 3898). They are much less common in developed countries, but there are some exceptions. For example in the United States, housing vouchers were randomly assigned to particular low-income households in the United States’ “Moving to Opportunity” programme (e.g. Kling, Liebman and Katz, 2007) or students to classes of different sizes in the “Project STAR” implemented in Tennessee (e.g. Hanushek, 1999).¹⁶ Particularly in the field of education there have been a relatively large number of experiments over the past decades (Fryer, 2016). The lower prevalence may be due to the perception that they are often considered politically difficult to implement or forbiddingly expensive. It may also be due to absence of influential institutions (e.g. development banks and agencies) promoting their use.¹⁷ Once treatment and control groups have been identified and the policy has come into effect, a simple comparison of average outcomes between the two groups can suffice to estimate the impact of the policy. This is the case if two conditions are met. First, random assignment ensured that the selection into treatment or control group is independent of the participants’ potential benefits from a policy. Second, the participants received comparable treatments, i.e. the same policy applies to all participants and the outcome of any participant is not affecting those outcomes of other participants.¹⁸ The latter assumption is particularly problematic when policies apply to relatively large geographical units or to people, firms or smaller spatial units within a given area. For example, if the opening of an enterprise zone in one (treated) municipality results in the relocation of a firm from another municipality that is part of the control group, the latter assumption would be violated.

Assuming that randomisation was successful and spillovers can be ruled out, one of two causal effects can be identified. The average treatment effect (ATE) or the average treatment effect on the treated (ATET). The difference might seem semantic, but can be substantial. The average treatment effect identifies the average causal effect for the whole population, i.e. the expected impact of the policy for a randomly selected person, firm or area out of the pool of all people, firms and areas. The average treatment effect on the treated captures the average causal effect only for those participants who actually received the treatment. If participants are randomly chosen from the population, the two treatment effects coincide, but if the pool of participants (both control and treated) is pre-selected, e.g. by certain criteria (only municipalities with a high unemployment rate, firms that apply for governmental support, etc.) the estimated effect is only valid for this selected subpopulation.

Regional policies have the potential to create ideal framework conditions for an RCT assessment of their impact. Impact assessment can (and should be) considered from the planning stage and eligibility for the policy can be made contingent on the willingness to participate in the assessment. The resulting projects can be implemented as “pilots”, meant to showcase and test the viability of a policy. Alternatively,

a policy can be rolled-out in a staggered manner, with the timing of the implementation being randomly assigned across participants.

While RCTs are the best way to estimate a causal impact, they have important limitations. First, the estimated impact applies to a specific treatment, e.g. a certain amount of funds, a certain type of industrial cluster, a specific amount of tax benefits for firms locating in a region, etc. The RCT does not indicate what happens if the amount of benefits is doubled or the tax benefits halved, or what would be the impact of an industrial cluster if additional infrastructure (e.g. a university campus) was introduced – unless this explicitly considered in the experiment design and part of the treatment. The second limitation is the scale of a policy. If a policy is successful with a limited number of participants, it might not necessarily lead to the same level of success when it is applied universally. An investment support programme in some municipalities might attract firms from other areas, but if the programme was available in all municipalities the incentive to relocate vanishes. A further concern is the “external validity”, i.e. whether the estimated impact can be applied in other contexts. Do, for example, results found in the United States also apply in the United Kingdom? Do they also apply in Eastern Europe or outside the OECD? Ultimately these worries can only be alleviated by replication. Finally, Deaton and Cartwright (2016) recently argued that the characteristic of RCTs of requiring minimal assumptions and prior knowledge is a clear advantage when “persuading distrustful audiences”, but is a crucial disadvantage for scientific research based on cumulative progress build on a wealth of previous contributions.

Quasi-random assignment: getting close to the “gold standard”.

As most policies are implemented without a RCT component, alternative methods to identify causal effects are required. Luckily, significant progress has been made in developing, adapting and understanding econometric techniques that address exactly this point. The methods aim at identifying causal relationships for policies that were implemented without random assignment into treated and control groups. Common to all these methods is the aim to create an (implicit) control group that mimics the attributes of an ex ante assigned control group.

The quasi-randomised methods can be distinguished by the mechanisms they exploit and classified into three broad groups: data driven, institution driven and mechanism driven methods. The first type of method relies on observable (and sometimes unobservable) characteristics to select a control group. The second type exploits that policies apply to specific people, firms or areas up to a certain (sharp) cut-off value, such as the age of a person, the size of a firm or the economic characteristics of an area. The third type of methods exploits the mechanism that determines the specific implementation or design of a policy. For example, the development of cities is often constrained by natural boundaries (mountains, lakes, sea) or new investments follow historic paths. The following gives a brief idea how these methods can and have been used, Baum-Snow and Ferreira (2015) provide an excellent review of the techniques discussed in this section and provide technical details.

The rationale for *data driven methods* is that differences between treated and untreated people, firms or areas are solely related to observable (and possibly some specific unobservable) characteristics. If this is the case, the characteristics can be taken into account and the assignment into either treatment or control (conditional on their characteristics) is as good as random. These types of methods are by far the most commonly used in empirical studies. They are also the most heavily criticised. The assumption that all relevant characteristics have been taken into account is untestable and the validity of the methods therefore relies on the believability of the assertion that all confounding factors have been accounted for. Dehejia and Wahba (1999) showed that this can actually be achieved. They use a US workforce training programme as example and compare RCT-based evidence with estimates obtained by using large scale survey data. Based on the characteristics of workers who were trained, they select comparable workers who did not receive training. The selection is based on age, education, marital status, gender, ethnicity and

(crucially) the past earnings of workers. In the estimates that control for all these factors the authors were able to replicate the RCT estimates within reasonable margins of error.

For regional policies, data driven methods face an additional challenge if policies apply at the regional level. While it is typically possible to draw on relatively large samples when assessing the impact of policies on people or firms, assessments at the regional level tend to have a significantly smaller number of (un)affected regions. This can result in situations where there are no natural control groups for regions that received a treatment, i.e. regions with similar characteristics. In this setting “synthetic control groups” have been suggested (Abadie and Gardeazabal, 2003). The idea is that no comparable regions exist, but that a (synthetic) composite of regions might proxy as a comparison. It is, however, important to consider the role of spatial spillovers before applying these methods (Gobillon and Magnac, 2016).

The second type of methods uses *institutional* factors to identify a control group. The main idea is that policies only apply to a group of people, firms or areas. Essential is that the selection is explicitly or implicitly based on some indicator with a certain cut-off distinguishing between those that are affected by a policy and those that are not. If there are enough observations of people, firms or regions around the threshold value, the difference between the two groups can be used to identify the impact of the policy. The key assumption is that observations around the threshold are very similar and that random factors determine their position just above or below the threshold. For example, Becker, Egger and von Ehrlich (2010) use the eligibility rule for EU (Objective 1) Structural Funds. Only regions with a per capita GDP level below 75% of the EU average were eligible for Objective 1 funds. Other examples for institutional methods are winner-takes-all situations, where the chosen person, firm or area and the runner-up(s) might form natural comparison groups. Greenstone, Hornbeck and Moretti (2010) use this design to assess the impact of large plant openings on other firms in the same county in the United States. They rely on the fact that these plant openings follow a long evaluation process and that the information on the chosen, as well as the runner-up counties was published. The assertion that the “winner” and the “runner-up” counties are comparable is plausible and further supported by a comparison of the characteristics of counties in the year prior to the plant opening.

The third type of methods aims to purge confounding factors from the evaluation by identifying exogenous information that isolates part of the *mechanism* that leads to a person, firm or area being affected by a policy. The idea is that a policy is partly determined by factors that are not directly related to the assessed outcomes. These factors are usually very policy specific and need to be justified on a case-by-case basis. To give a concrete example, infrastructure investment, such as major roads, is not randomly allocated but its location is carefully planned. But today’s investment is influenced by historic paths that were often constructed for very different reasons (e.g. to transport military or as part of a grid structure). This suggests that historic infrastructure could be used to identify exogenous variation in today’s network, if it can be argued that the historic choices are not determined by the same (unobservable) factors that relate to today’s outcomes. Studies for the United States (e.g. Michaels, 2008), India (Donaldson, forthcoming) or the People’s Republic of China (Baum-Snow et al., 2015) have used historic road or railroad plans to causally identify the impact of modern road or rail infrastructure on varying outcomes.

5. Scale and spillovers: Challenges specific to the evaluation of “regional” policies

As Gibbons, Nathan and Overman (2015) clearly point out, “Research analysing spatial data often falls short of the standards set by current research in other policy areas – for example in clinical treatments and aspects of active labour market policy or international development.” The reasons for that are manifold, as the same authors argue. First, there are some specific data limitations, as well as methodological and conceptual challenges that make the analysis of “spatial” data more difficult than other types of data. Second, in urban and regional economics the adoption of econometric tools for counterfactual impact evaluation has been slower than in other fields (e.g. labour or health economics).

This is partly due to the conceptual and methodological challenges mentioned in the first point, and partly to the co-existence in the discipline of different views on the degree of applicability of quantitative economic methods to urban and regional analysis. Third, as already discussed in this report, the definition of the policy objectives of regional or urban policies is often more blurred than it is the case with other kinds of policy interventions. Therefore, in absence of clear and measurable objectives, the evaluation exercise becomes a much more difficult endeavour.

This section discusses in a non-technical manner the first issue, i.e., the methodological and conceptual challenges that specifically affect the evaluation of policies that have a subnational dimension. The second point involves epistemological issues which are beyond the scope of this paper; the interested reader can find in Overman (2004) an example of this kind of debate. The third point is discussed in the next Section of this report.

The choice of the right spatial unit and of the right data

Economic analysis at subnational level has always been hampered by the scarcity of good data sources and by arbitrariness and conceptual complexity of the definition of subnational spatial units. Data scarcity is a rather simple issue: the collection of statistics at subnational level is obviously more burdensome and less precise than the equivalent task at national level. Confidentiality rules are generally more binding due to the smaller reference population.

The methodological problems are more subtle. While at national level the definition of the border of the geographical unit is mostly uncontroversial, at subnational level the researcher is faced with a number of different alternatives (e.g. regions, provinces, municipalities, travel-to-work areas, local labour markets, functional urban regions, etc.), each of them bearing different properties, information breadth, and sample size. Using a spatial unit that does not fully fit the economic phenomenon under scrutiny is not trivial and it is related to the so-called "modifiable areal unit problem" (MAUP). The MAUP arises from the large and unpredictable variation that economic estimates may have depending on the size and shape of the adopted spatial units. Every geographical area may be divided in many different ways, and commonly used statistics may present huge variation among them. Furthermore, differently from international comparison (where country borders have an economic and political meaning that is not comparable with any other geographic classification) in a subnational setting researchers face a variety of administrative and functional divisions – each of them with its pros and cons – with the result that the choice of the spatial unit is often arbitrary, or constrained by data availability. The MAUP has been widely investigated and debated. The first contributions go back to as early as the 1930's with more recent work by Openshaw (1984), Arbia (1989), and many others.

This is why using administrative units – like provinces or regions – is the simplest, and most common, solution, but often not the most suitable one. In most cases administrative borders are a historical legacy and have progressively lost economic meaning over time, especially in Europe. This means that generally urban agglomerations or local labour markets cross municipal, provincial and regional borders. The metropolitan area of Paris is probably the most striking example. The agglomeration extends well beyond the city of Paris or even the Île-de-France region and covers a total of 1 375 municipalities in the Île-de-France and four adjacent regions. At the same time, administrative units like provinces and regions typically contain several urban agglomerations and local labour markets of different sizes with different economic fundamentals. In the United States the situation is similar, with county boundaries bearing very little economic significance nowadays, and states being too large for most local economic analyses. In addition, counties vary significantly in size between the east and the west of the country. This is however partially compensated by the availability of statistics from the US Census at the level of metropolitan statistical areas (MSAs), defined as aggregations of counties linked by strong commuting flows. In Europe, harmonisation of definitions has taken some time, despite the seminal efforts of Cheshire and Hay (1989).

But with the development of the harmonised definition of “Functional Urban Areas” for EU and OECD countries (OECD, 2012) aggregate statistics and time series have been made available for metropolitan areas and for some urban agglomeration from official data providers like Eurostat and the OECD.

A possible way for researchers to overcome the inherent limitations of administrative subnational data sources consists in being creative – i.e., in creating new databases from innovative sources of data at “micro” level. The collection, visualisation, storing, elaboration of micro(geographic) data has been greatly facilitated by the progress of computer science, and new datasets and geographic information system (GIS) platforms offer many valuable opportunities for researchers.¹⁹ Therefore, there is increasingly more opportunity to use spatial units that are the most meaningful in relation to the economic phenomenon under scrutiny or to dispense with discrete spatial classifications altogether, adopting a continuous definition of space e.g. using firms’ addresses (Duranton and Overman, 2005). Other seminal examples of creative use of new microdata sources in spatial economics are the analysis of satellite land-use data (Burchfield et al., 2006) or of nightlight data (Henderson, Storeygard, and Weil, 2012). An increasing number of data sources contain now detailed geographical information (e.g. postal code or GPS coordinates) which can be easily allocated to administrative or functional spatial units.

However, targeted or ad-hoc definitions of spatial units might have some shortcomings for policy analysis. As Gibbons, Nathan and Overman (2014) point out, often policy makers are interested in knowing how “their” city or region is performing. Furthermore, the spatial scale at which future policies will be implemented will still be the administrative one. Therefore, it might still be informative to perform the analysis (also) at the administrative level while properly taking into account that some of the effects of the policy may either spill over to neighbouring regions, or be limited to parts of the area under scrutiny. This in turn means that the evaluation of local policies could also inform policy makers on the most appropriate administrative level at which the policies should be implemented.

The risk of a zero-sum game: Spatial substitution effects and sorting

The positive effects of a regional or urban policy on eligible areas may come at the cost of a specular negative effect in ineligible areas. One common mechanism through which this could happen is sorting. As people and firms are generally free to choose their location, the observable effects of a programme may be due to potential beneficiaries selecting to move to supported localities. Whether this can lead to misinterpreting the results of the evaluation exercise depends on the policy design and objectives. For instance, sometimes the policy is exactly aimed at attracting firms and investments: in this case, the sorting of firms with specific characteristics into the target areas is the expected outcome. In other cases, however, sorting can be an undesirable or neutral outcome: this could be the case, for example, of a regional R&D support scheme that simply leads to the relocation of R&D performing firms within the country, with no additional effect on the aggregate amount of R&D investments nationwide; or of a new motorway improving accessibility to a particular municipality may lead all firms in the neighbouring municipalities to concentrate around the access point, with zero net effects on total employment across the whole wider region.

In the tradable sector, spatial substitution effects can also materialise in the absence of sorting of people and firms. The mechanism is similar, but instead of relocating firms or people, production sorts to the most profitable locations. The positive effect in a region comes at cost of a negative effect in its neighbours (“crowding out”). This may happen, for example, in cases in which neighbouring similar areas compete in the same markets for the same final demand. A policy that sets spatial boundaries gives firms in the area where it applies an edge over its competitors, resulting in increased market shares and output. The crowding out can be particular relevant for the tourism sector, e.g. development of a new attraction might not create new tourism but simply shift it away from existing attractions.

To assess whether a policy has substitution effects, the impact of a policy needs to be estimated at different spatial scales. Without substitution, the impact at larger spatial scales should be the same as at lower spatial scales, e.g. if a policy is estimated to create 1 000 jobs at the municipal level, it should also create 1 000 jobs at the provincial level. Of course this comparison only holds when absolute counts are considered and not percentage changes or ratios, e.g. the unemployment rates. A degree of substitution is not necessarily an undesirable outcome, if one of the stated policy objectives is to reduce spatial inequalities, but it might not be the most cost-effective way to reach such an objective. In some circumstances, sorting may also lead to a negative effect of the policy, for instance in the case in which the subsidy pushes rent-seeking companies to choose a sub-optimal location where their productivity is lower.

Accounting for the first law of geography: Spatial autocorrelation and spatial spillovers

The discussion of substitution effects is useful to highlight that spatial units are not isolated entities, on the contrary, they often show a high degree of economic interaction with neighbouring spatial units (see e.g. the discussion by Gordon, 2001, related to local labour markets). This is at the base of the rather simple “first law of geography”: “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970, 236). In empirical analysis at subnational level, this translates into the so-called spatial autocorrelation, i.e., a significant statistical association of values across neighbouring units. In turn this may be due to correlation in unobserved characteristics, to common “contextual” factors, and/or to spillover effects.

Distinguishing between spatial autocorrelation arising from common unobserved characteristics, on the one hand, and from spatial spillovers, on the other hand, is difficult. The fact that nearby agents share common outcomes or characteristics may depend on a number of underlying factors. Manski (1993) distinguishes three reasons why the outcome variable may present correlated values across members of a same group – or spatially contiguous observations in the context of this paper. In Manski’s taxonomy the link is either “endogenous”, due to “contextual factors”, or “correlated effects”. It is easiest to consider the distinction in a concrete example, e.g. a transport infrastructure programme (the policy, e.g. a new motorway) that is assumed to positively affect employment (the outcome) in a region.

First, employment in a region may directly affect employment in other regions: this is the “endogenous” effect in Manski’s taxonomy. In the chosen example, it would imply that an increase in employment in a region affects employment also in neighbouring regions, e.g. through higher demand spurred by input-output linkages across firms in the contiguous regions. A second mechanism operates through the policy, which may present itself some degree of spatial autocorrelation: this is the “contextual factor” in Manski’s taxonomy. E.g. the new motorway crosses a group of neighbouring regions which therefore share a similar “treatment” – and indirectly also a similar employment effect. Third, there are “correlated effects”, which are common to all regions and also affect employment, but are unobserved to the researcher: e.g. a group of contiguous regions might have a higher endowment of human or social capital, which, by chance, happens to be concomitant with the new motorway. Importantly, only the first effect can properly be labelled as “spillover effects”. However, all the three effects translate into spatial autocorrelation in the outcome variable, which means they are observationally equivalent, resulting in Manski’s well-known “reflection problem”. Some specific econometric techniques are available to try to disentangle them (e.g. see Anselin et al. 2013), but these typically require strong assumptions on the structure of the spillovers that are derived from economic theory and tests to the validity of these assumptions remain often elusive.

The economic literature identifies several different kinds of spatial spillovers: knowledge spillovers materialise when the knowledge created by a firm or a research institution also benefits other neighbouring entities; sectoral spillovers spur from input/output or competition linkages that create “domino” effects on (horizontally or vertically) related industries; growth spillovers, more generally, refer to the situation in

which the economic dynamics of a spatial unit influence its neighbours through trade linkages and market relationships (Capello, 2009).

Translated into a policy evaluation context, spillover effects consist of the effects of the policy not bound within the observed spatial unit and affecting also neighbouring areas. In the evaluation phase, spatial spillovers are deeply interrelated with the choice of the right spatial unit, and more generally with the adoption of the appropriate empirical methodology able to tackle this additional component. Spatial econometrics provides some solutions to explicitly account for this spatial dependence by using a “spatial lag” or a “spatially autocorrelated” regression model, with the choice depending on the channel or indicator through which the spatial dependence arises (Anselin, 1988), though analysis should keep the first order priority of establishing a causal link (as outlined in the above section) in mind (see also Gibbons and Overman, 2012). A typical example for spatial autocorrelation of policies is the analysis of local taxation dynamics, as often neighbouring municipalities “compete” in attracting businesses or people and therefore the choice of a tax rate for a given municipality directly affects also neighbouring municipalities’ choice of tax rates (Brühlhart, Bucovetsky and Schmidheiny, 2015).

Spatial spillovers belong to the category of pure externalities, i.e., they produce uncompensated advantages for recipients, leading to a discrepancy between private and social optimum (Capello, 2009). Therefore, the existence of spatial spillovers may justify policy interventions to correct a market failure. It follows that understanding how spillovers work is equally important in the policy design phase as much as it is in the evaluation phase, and it should be taken into consideration for the choice of policy levers, the spatial scale of policies and their implementation.²⁰

Not everything can be measured: Local unobservable factors

Overcoming the potentially confounding impact of unobservable characteristics is an important challenge in any evaluation exercise. Unobserved characteristics confound the analysis if two conditions are met. First, the unobserved characteristics determine whether a person, region or firm was selected as beneficiary of a policy and second, the characteristics also determine the outcome of interest. If both conditions are met, an estimate that does not account for the unobserved characteristics will be biased as it partially attributes the impact of the unobserved characteristics to the policy. For subnational analyses this potential bias originates in the failure of take into account relevant local unobservable factors. In Manski’s terminology, this means that “correlated effects”, e.g. a higher level of human capital in regions that were selected to benefit from a new motorway, are not taken into account. This point may just appear as a specific case of the well-known “omitted variable bias”. But there are some important differences when working with subnational data, compared to e.g. cross-country or individual-level analyses. The main difference is that there are many factors that are likely to be relevant for economic growth in regions and cities and that their dynamism is often difficult to conceptualise and quantify for the researchers, in part due to a lack of reliable data. Examples of those factors include social capital and trust (e.g., Guiso, Sapienza, and Zingales, 2004), “civicness” (de Blasio and Nuzzo, 2010), cultural values and attitudes, know-how, openness to innovation (Florida, 2002), etc. This is relevant in the context of evaluation because these factors may explain why an apparently “good” counterfactual may still lead to biased results.

As a further cause of complexity, all these factors tend to vary smoothly over space, and administrative borders might have little or no correspondence to real spatial discontinuities; therefore, they can also contribute to exacerbate spatial autocorrelation in the outcome or treatment variables (via a spatial correlation structure in the error term in the regression model). For instance, it is often argued that the level of civic-mindedness across Italian regions and municipalities it is strongly related to historical factors like foreign dominations, with political geography from the Middle Ages or the Renaissance bearing more explanatory power than the current one (Guiso, Sapienza, and Zingales, 2010).

All in all, for the researcher this implies that the list of potentially unobserved variables is even longer, and more likely to introduce a non-trivial bias, than in cross-country or individual-level analyses. In an evaluation exercise, this makes the choice of the control group more complicated, as apparently comparable spatial units may substantially differ in terms of those unobservable factors. In light of that, the development of a solid identification strategy possibly exploiting exogenous variation in the treatment variable acquires further importance.

6. Good practice: Some recent benchmark evaluations

This section provides a few examples of impact evaluation of policy programmes with a subnational component that are linked to the issues discussed in the report. The survey does not aim at being exhaustive and the interested reader is referred to the summaries in different thematic areas provided by the UK What Works Centre for Local Economic Growth.²¹ This brief survey, therefore, covers selected academic contributions that are easily linkable to the other points discussed in the report and that provide generalizable examples.

Transport infrastructure

Transport infrastructure investment follows (expected) demand. A positive relationship between new roads or train connections and local development could therefore simply reflect positive prior trends or other confounding factors. Recent studies have used historic road or rail development plans to try to eliminate confounding factors. This strategy mixes mechanism and institutional factors and relies on historic planning decisions being independent from contemporaneous economic or demographic developments.²² Several authors have used the 1947 US plan of the interstate highway system to identify exogenous variation in available transport infrastructure (Baum-Snow, 2007; Michaels, 2008; Duranton and Turner, 2012). The plan aimed to link cities, industrial centres and suitable border points and support national defence (see Michaels, 2008 for details). But the plan inadvertently also affected other local transport conditions, which is what the different papers exploit.

Baum-Snow (2007) estimates the impact of highways provision on suburbanisation instrumenting the number of “rays” of highways extruding from a core city with the number that would have been present if the 1947 plan would have been implemented by distributing funds uniformly to construction within the state. Michaels (2008) focuses on rural counties and estimates the impact of connecting a county to the interstate highway system on employment in different sectors. He instruments the presence of a highway with a planned route crossing the county. He introduces an additional instrument, using that the plan follows a grid pattern that connects major cities. If a rural county has its nearest city located exactly north, east, south or west the likelihood that it would be crossed by a highway is higher than for counties with the nearest city located at an angle. Duranton and Turner (2012) consider the impact of highway infrastructure on employment and population growth in US metropolitan statistical areas. They instrument the stock of highway kilometres with the planned stock of highway kilometres. In addition they use two further instruments based on historical transport patterns that can explain modern transport networks because they indicate less costly routes for highway development. The first instrument is the length of railroad in 1898, the second is based on the kilometres of exploration routes of major expeditions between 1528 and 1850.

More recently, Ghani, Goswami and Kerr (2015) investigate the impact of transport infrastructure on the organisation and efficiency of manufacturing activity in India. The project under scrutiny is called Golden Quadrilateral (GQ), and it consisted in upgrading a central highway network. The authors find that manufacturing activity grew disproportionately along the network. Both entrants and incumbents facilitated the output growth, with scaling among entrants being important. Some of the positive outcome might, however, depend on the fact that the most successful Indian districts managed to get connected to the new highway, while the lagging districts did not. If this was the case, some of the positive effect

attributed to the highway would be spurious. To ensure that their results are indeed causal, the authors adopt two independent strategies. The first is based on constructing a hypothetical highway network that connects the metropolitan areas that form the corner points for the highway via straight lines.²³ The second approach is based on a falsification exercise: the impact of GQ is compared to the estimated impact of a second highway that was planned to be upgraded at the same time as GQ but subsequently delayed. The authors also show that the estimates that do not take into account the endogeneity issue underestimate the true impact for wage and productivity. This therefore suggests that “underperforming” firms were more likely to get connected by the GQ infrastructure.

An empirical paper that is interesting for its close fit to the theoretical predictions of the New Economic Geography is Teixeira (2006). The paper assesses the impact of transport investments in Portugal over the period 1985 to 1998. The empirical test is based on a structural model²⁴, which is closely derived from the models derived in the “New Economic Geography” literature. The model replicates the observed patterns in the data well. This partially compensates for the lack of an explicit counterfactual strategy in the analysis. The author finds that Portuguese transport policy has not contributed towards spatial equity, although the model predicts that further investments would have lowered transport costs enough to lead to a wider spread of industrial production. This is consistent with a bell-shaped relationship between transport costs and agglomeration: for high transport costs, an initial reduction leads economic activity to further concentrate in the “core”, while further reductions lead to a spread of industrial production to the periphery.

Policies for lagging regions

Criscuolo et al. (2016) exploit changes in the area-specific eligibility criteria of the “Regional Selective Assistance” programme in the United Kingdom. The programme aimed to support manufacturing jobs. The eligibility changes arise from modifications of the EU regulations on state aid every seven years. As it is plausible that changes in the EU regulations are independent from local unobservable factors in the United Kingdom, the change in eligibility criteria lends itself as an “institutional” identification strategy. The authors use the predicted change in eligibility based on the EU eligibility rule and characteristics of the area that predate the period under study. The authors find positive programme treatment effect on manufacturing employment without evidence of displacement. The effect comes solely from jobs created in smaller firms below 150 employees, whereas large firms seem to “game” the programme. Estimates that do not tackle the endogeneity problem underestimate the effects of the programme, because the policy targets underperforming plants and areas.

Becker, Egger, and Ehrlich (2012) assess the impact EU funds provided to “Objective 1” regions in the period 1989-2006. Objective 1 regions are those with a per capita GDP level below 75% of the EU average in the pre-funding period. The 75% GDP threshold allows the authors to implement a regression-discontinuity design that exploits the discrete jump in the probability of being a recipient of the EU transfer for identification of causal effects. This institution-driven design is based on the assumption that regions that are just above the 75% threshold are a valid control group for regions that are just below, and therefore differences in economic outcome are attributable entirely to the effect of the policy.²⁵ The results show positive per capita GDP growth effects, but no employment growth effects.

Policies to attract investments

Greenstone, Hornbeck, and Moretti (2010) examine the effects of large plant openings in the United States on incumbent plants. These “million dollar plants” are often subsidised by local policy makers keen to attract new investments to their constituency. In order to base their estimation on plausible control group, the authors use the location rankings of profit-maximising firms. Incumbent plants in the county where the new plant ultimately chose to locate (the “winning county”) are compared to incumbent plants in

the runner-up county (the “losing county”). To substantiate their identification strategy, they show that incumbent plants in “winning” and “losing” counties have similar trends in total factor productivity in the seven years before the new plant opening. Five years after the new plant opening, productivity of incumbent plants in “winning” counties is 12% higher than in incumbent plants in “losing” counties. In line with urban economic theory, the effect is larger for incumbent plants that share similar labour and technology pools with the new plant. Benefits also accrue to workers as the results also point to a relative increase in wages in “winning” counties, a finding that is in line with general equilibrium theories of local labour markets.

7. Conclusions and recommendations

This report reviews both the main theoretical and empirical aspects of policy design and evaluation in the field of policies that have a subnational dimension (regional, urban or “local” policies). Building on prior OECD work (including notably Warwick and Nolan, 2014; and Oldsman, 2014) it aims at reducing the perceived distance between policy makers and economists in the field.

The advantages of rigorous impact evaluation go beyond the crucially important need of learning about which programmes work, and which do not. Good evaluation can help the policy maker to insulate from pressure groups and vested interests; it can inform on the timing of the policy effects; it can uncover unintended or unexpected positive or negative effects of a programme. As often stressed in this report, impact evaluation is also closely intertwined with policy design, and improvements in both phases of the policy cycle are self-reinforcing.

This report also argues that in the next years OECD member countries will be confronted with the demand for more effective and efficient regional and urban policies. These policies have to be more effective because the recent economic trends have been pushing toward an increase in spatial socioeconomic disparities within countries, leading to a “Great Divergence” among lagging and leading agglomerations, as Enrico Moretti (2012) labelled it. Policies with a subnational focus (but also economic policy in general) also have to be more efficient because public sector budgets are under pressure in many OECD countries, and there is increasing expectation for value for money in policy making.

But rigorous evaluations remain scarce. The researchers from the “What Works Centre for Local Economic Growth” (WWG) reviewed around 2 300 impact assessments in the field of transport investments, to find out that only 29 are based on a sound methodology that can properly estimate *causal* effects; the figures are similar in the fields of access to finance, apprenticeships, business advice, etc.

But there remains a certain reluctance to commit to rigorous policy evaluation. In part this is an issue of political economy. Once a policy has been implemented, vested interest have been created that could suffer if the results of an evaluation are not positive. Ensuring a commitment to evaluation before the policy is enacted and communicating clear adjustment mechanisms might help address this challenge. Another challenge is that in many cases evaluations do not find large or even positive effects. Only two of the 29 studies the WWG considered for transport investment find a positive outcome. While it is uncontroversial that some programmes may fail to reach the targeted objectives, the overwhelmingly negative results may also be a reflection of limited or poor data availability and of the academic standard to set a high threshold to distinguish between statistical noise and an actual correlation between a policy and an outcome of interest. Therefore, there is also the need for more and better, but also for more policy experimentation in the field. However, experimentation is closely intertwined with evaluation, as experimentation is of little use if its results are not properly evaluated.²⁶

The report also points to a number of concrete actions that policy makers can take in order to get the most from evaluation of regional or urban policies. These messages also reflect the conclusions by Gibbons, Nathan, and Overman (2014) and the reviews of the WWG. The interested reader may also refer to Oldsman (2004) for a complementary set of policy recommendation in the field of innovation policy.

- **Clarify policy objectives and levers.** As compared to other policy areas, “local” economic policy is characterised by weaker evidence on which policy instruments actually work, and sometimes also by a lack of clarity on the policy objectives. In order for the policy to be properly implemented and evaluated, it is of paramount importance that both the intended outcomes and the policy instruments are made explicit.
- **Choose the right spatial level for design, implementation, and evaluation.** Those should not necessarily coincide. In particular, in the evaluation phase it is important to take into account the linkages and the spillovers across neighbouring areas, and the fact that often the choice of the spatial unit of analysis is not neutral to the results of the empirical analysis. I.e., the measured impact of a policy at a given spatial scale can be substantially different at a different spatial scale (e.g., municipalities versus regions).
- **Develop a theoretical understanding of the mechanisms at play** in order to take into account possible counterintuitive outcomes or unintended consequences of the programme in the design phase. The “New Economic Geography” literature, for instance, gives a number of useful predictions which have been progressively empirically validated over the last two decades. Conversely, some popular policies may be grounded on weak theoretical basis (e.g. cluster policy) or may have an expected theoretical impact which is at odd with the stated policy objective (e.g., transport investments in peripheral regions leading to further agglomeration in the core regions).
- **Make the evaluation culture pervasive** through all stages of policy design and implementation (and reform), e.g. through targeted training and partnership with independent evaluation agencies and academic institutions. Often it is easier to elicit “buy in” into an ex-post evaluation of a policy before it is implemented and funded, as ex-post vested interests might make an assessment more difficult. Successfully implementing an evaluation culture also requires space for experimentation and failure. Not all policies work. Rewarding success only can lead to reluctance in trying new ideas and a reliance solely on policies that have worked in the past. But this means that the evaluation will no longer be able to answer the question “what works?”, but only “what has worked?” and potentially misses out on innovative and successful new ideas.
- **Learn from other evaluation analysis** and disseminate the results of “your” evaluations. Once the main conceptual and methodological aspects are properly mastered by the policy makers, accessing a wide knowledge base of evaluation exercises – with the ability to discern higher and lower quality analyses – becomes a critical asset in the phase of policy design. There is also room for organizations like the OECD in acting as a facilitator for the sharing of international experiences in evaluation practices.
- **Embed evaluation and data collection in policy design**, and support an “open data” policy throughout all the stages and levels of policy implementation. This would increase the accountability of evaluation, as the analysis would be easily replicable. It would also make evaluation available “for free” to the policy community, as publishing incentives could leverage the academic expertise in the field.²⁷

- **Bring the evaluation results back to the design phase.** A good evaluation exercise should make explicit what the results imply for the new versions of the programme. Ideally, the possibility that the programme should be discontinued in case of no observable positive outcome should not be disregarded. More realistically, the evaluation report should underline all the possible marginal improvements that would lead to a better programme.
- **Take a holistic approach and learn from the beneficiaries and the programme managers.** It is not rare that a programme succeeds or fails to achieve its intended outcome because of unexpected or unintended factors. Not all aspects that translate a policy from inputs and outputs to outcomes are under policy makers' control. More generally, the technicality of the functioning of the programme, complex behavioural components affecting the programme's outcome or the selection of beneficiaries, and other details that are never considered in a stylised economics textbook may play a very important role. Also, it is useful to remember that a counterfactual evaluation, while being a powerful tool, is only one tool in understanding the impact of policies and should not be insulated from the consideration of other "softer" and qualitative aspects of the programme under scrutiny. However, qualitative research should follow the same methodological rigour that is advocated in this report for quantitative research – for instance in terms of the consideration of a counterfactual scenario and of the careful identification of causal links.

NOTES

- 1 . A number of governments are actively pursuing strategies to make rigorous policy evaluation part of their decision process. The United Kingdom, for example, has created the “What Works Centre for Local Economic Growth” with the mission to analyse which policies are most effective in supporting and increasing local economic growth. The Directorate General for Competition of the European Commission has recently reformed its State Aid rules strengthening the role of evaluation (“State aid control 2.0”) (European Commission, 2014a).
- 2 . The link between experimentation, failure, and evaluation, however, can also create a vicious circle in which the fear of a negative outcome from evaluation provides a disincentive to experimentation, of which failure is an essential component.
- 3 . For example, OECD (2015a) highlights the growing prevalence and rigour of ex-post regulatory impact assessments among OECD countries. While in the early-1990s less than 10 OECD countries had adopted regulatory impact assessments, by 2014 all but one OECD country used the tool systematically.
- 4 . Breinlich, Ottaviano and Temple (2014) give the examples of the Italian and English North-South divides. See also OECD (2016a), OECD (2016b) and OECD (2013) for further evidence from OECD countries.
- 5 . The OECD classifies regions into two territorial levels that reflect the administrative organisation of countries. OECD’s large regions (TL2) represent the first administrative tier of subnational government, such as the Ontario region in Canada. OECD small (TL3) regions are contained within a TL2 region. For example, the TL2 region of Castilla-La Mancha in Spain encompasses five TL3 regions: Ciudad Real, Cuenca, Guadalajara, Toledo and Albacete. In most cases, TL3 regions correspond to administrative regions, with the exception of Australia, Canada, Germany and the United States.
- 6 . Evidence for large “NUTS 2” regions according to the EU nomenclature of territorial units for statistics (NUTS).
- 7 . Important steps are, however, currently being taken to instil an “evaluation culture”. For example, the EU Structural Funds regulations have included evaluation requirements for several programming periods and most recently have increased their focus on the impact of the interventions. According to the general regulation for the 2014-20 period, the effectiveness, efficiency and impact of assistance from the Funds will be evaluated in order to improve the quality of design and implementation of programmes (European Commission, 2014a).
- 8 . See the Policy Forum in OECD (2011).
- 9 . The growing economic importance of services has significant implications for the spatial distribution of economic activity. Goods manufacturing tends to be more space-intensive than the delivery of services and the benefits from co-location tend to arise through different channels. While clusters in manufacturing often formed around shared local factor inputs (e.g. woods, rivers, mines), benefits in services are thought to be more related to the diffusion of knowledge that is greatly facilitated by face-to-face interactions.
- 10 . The most recent developments of urban economic theory can explain other important stylised facts of contemporary urban geography, including: why earnings and housing rents are increasing in city size; why larger cities are more unequal (the Gini coefficient of urban earnings is positively correlated with city size); why urban productivity premium is increasing in the education level; and why the city size distribution follows closely a power law (Behrens and Robert-Nicoud, 2015).

- 11 . This outline focuses on agglomeration economies in producing goods, there are also agglomeration economies in consumption, i.e. larger cities offer a wider variety of services, greater differentiation in the quality of goods and a certain size might be required to fund certain public or private services (e.g. operas, top-flight sports clubs, etc.).
- 12 . Another example of a very insightful contribution is the recent paper by Kerr and Kominers (2015), exploring how different agglomeration forces lead to different cluster shapes. The authors build a model of clusters comprised of small, overlapping regions, to show that agglomerative forces with longer spillover distances are associated with fewer, larger, and less dense clusters. Their model also helps explain why innovation clusters can be concentrated in small geographic regions.
- 13 . The classification of input, output and outcome indicators as monitoring “effort”, “efficiency” and “effectiveness” follows Schumann (2016).
- 14 . Robert Frost (1916) “The Road Not Taken”, available at <http://www.poets.org/poetsorg/poem/road-not-taken> (accessed 2 April 2015).
- 15 . More technically, the treated group or even individual person, firm or area is compared with a group, person, firm or area that is similar “in expectations” or for which the difference disappears in the (probability) limit, i.e. as the number of observations increases to infinity. This means that even when a single treated unit is considered, the estimated effect is still not an “individual” causal effect but rather the average causal effect.
- 16 . The provided examples are some of the more prominent RCTs with a spatial dimension, but their implementation is far from perfect compared to the standards that are expected in RCTs in medical drug trials. One of the critical drawbacks in most RCTs that aim at economic development is that a “double blinding” of the trial is often impossible.
- 17 . Institutional status-quo seems to be the main cause of resistance to experimentation and RCTs in organizations: <http://www.innovationgrowthlab.org/blog/what-prevents-organisations-using-more-evidence-and-rcts> (accessed 22 August 2016)
- 18 . Rubin (1986) gives a brief discussion of the second assumption, called the “stable unit treatment value assumption” (SUTVA).
- 19 . For a detailed discussion of application of GIS in economics, see Overman (2008).
- 20 . These may not necessary coincide, as it could be more efficient to design a policy that covers a wide area, and to implement the policies at a smaller spatial scale. Ideally, the evaluation should then be implemented at both spatial scales. However, the discussion tends to become vague and in general terms and the correct choice should be assessed case-by-case.
- 21 . The WWG reviews are available at <http://www.whatworksgrowth.org/resources/>
- 22 . More specifically, historic decisions only affect current outcomes through their impact on the policy under study, i.e. in this case through the available transport infrastructure.
- 23 . See also Faber (2014), who uses a similar instrument based on the least-cost path that could be used to connect the main cities within the Chinese trunk highway network.
- 24 . A structural model is a system of equations, which is derived from economic theory and is used for empirical work through the calibration of some parameters of the model that are assumed to be invariant to changes in the economic environment.

- 25 . The authors' data show that some regions were (not) receiving funds even though they had per capita GDP above (below) the 75% threshold. A variety of reason (including ex post data adjustments) can explain these issues. To ensure that the regression discontinuity remains valid, they use a “fuzzy” approach that instruments observed Object 1 status with the predicted status given the eligibility rule and the authors' per capita GDP data.
- 26 . The activity of the Innovation Growth Lab (<http://www.innovationgrowthlab.org/>) is particularly interesting in this context. The Innovation Growth Lab (IGL) is a new global collaboration promoting pilot experiments in the field of innovation and entrepreneurship policy. The aim is “to enable, support, undertake and disseminate high impact research that uses randomised trials to improve the design of the programmes and institutions that help make economies more innovative and entrepreneurial”.
- 27 . This recommendation is also stressed in OECD (2015d) and supported by OECD (2015e), which highlights the potential benefits for the public sector of bolder open data policies.

REFERENCES

- Abadie, A. and J. Gardeazabal (2003), “The Economic Costs of Conflict: A Case Study of the Basque Country”, *The American Economic Review*, Vol. 93(1), pp. 113-132.
- Ahrend, R. et al. (2014), “What Makes Cities More Productive? Evidence on the Role of Urban Governance from Five OECD Countries”, OECD Regional Development Working Paper, 2014/05, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5jz432cf2d8p-en>.
- Ahrend, R. and A.C. Lembcke (2016), “Does It Pay to Live in Big(ger) Cities?: The Role of Agglomeration Benefits, Local Amenities, and Costs of Living”, OECD Regional Development Working Papers, No. 2016/09, OECD Publishing, Paris, <http://dx.doi.org/10.1787/e0490ba8-en>.
- Anselin, L., R. Florax and S.J. Rey (eds.) (2013), *Advances in spatial econometrics: methodology, tools and applications*, Springer Science & Business Media.
- Audretsch, D.B. and M.P. Feldman (1996), “R&D spillovers and the geography of innovation and production”, *The American Economic Review*, Vol. 86(3), pp. 630-640.
- Bartolini, D., H. Blöchliger and S. Stossberg (2016), “Fiscal Decentralisation and Regional Disparities”, *Economics Department Working Paper*, No. 1330, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jlpq7v3j237-en>.
- Baum-Snow, N. (2007), “Did Highways Cause Suburbanization?”, *The Quarterly Journal of Economics*, Vol. 122(2), pp. 775-805.
- Baum-Snow, N. and F. Ferreira (2015), “Causal Inference in Urban and Regional Economics”, in Duranton, G., J.V. Henderson and W.C. Strange, *Handbook of Regional and Urban Economics*, Vol. 5, Ch.1, pp. 3-68.
- Becker, S.O., P.H. Egger and M. von Ehrlich (2010), “Going NUTS: The effect of EU Structural Funds on regional performance”, *Journal of Public Economics*, Vol. 94(9-10), pp. 578-590.
- Behrens, K. and F. Robert-Nicoud (2015), “Agglomeration Theory with Heterogeneous Agents”, in Duranton, G., J.V. Henderson and W.C. Strange, *Handbook of Regional and Urban Economics*, Vol. 5, Ch. 4, pp. 171-245.
- de Blasio, G. and G. Nuzzo (2010), “Historical Traditions of Civicness and Local Economic Development”, *Journal of Regional Science*, Vol. 50(4), pp. 833-857.
- Breinlich, H., G.I.P. Ottaviano and J.R.W. Temple (2014), “Regional Growth and Regional Decline”, in Aghion, P. and S. Durlauf (eds.) *Handbook of Economic Growth*, Vol. 2B, Ch. 4, pp. 683-779
- Brodeur, A. et al. (2016), “Star Wars: The Empirics Strike Back”, *American Economic Journal: Applied Economics*, Vol. 8(1), pp. 1-32.

- Brühlhart M., S. Bucovetsky and K. Schmidheiny (2015), “Taxes in Cities: Interdependence, Asymmetry, and Agglomeration”, in Duranton, G., J.V. Henderson and W.C. Strange, *Handbook of Regional and Urban Economics*, Vol. 5, Ch.1, pp. 3-68.
- Burchfield, M. et al. (2006), “Causes of sprawl: A portrait from space”. *The Quarterly Journal of Economics*, 587-633.
- Capello, R. (2009), “Spatial spillovers and regional growth: a cognitive approach”, *European Planning Studies*, Vol. 17(5), pp. 639-658.
- Carlino, G. et al. (2012), “The agglomeration of R&D labs.” Federal Reserve Bank of Philadelphia Working Paper 12-22.
- Carlino, G. and W.R. Kerr, (2015), “Agglomeration and innovation”, in Duranton, G., J.V. Henderson and W.C. Strange, *Handbook of Regional and Urban Economics*, Vol. 5, Ch.1, pp. 3-68.
- Cheshire, P.C., M. Nathan and H.G. Overman (2014), *Urban economics and urban policy: Challenging conventional policy wisdom*, Edward Elgar Publishing.
- Cheshire, P. C., and S. Magrini (2006). “Population growth in European cities: weather matters—but only nationally”, *Regional Studies*, Vol. 40(1), pp. 23-37.
- Cheshire, P.C. and D.G. Hay (1989), *Urban problems in Western Europe: an economic analysis*. Hyman.
- Combes, P.P., and L. Gobillon (2015), “The Empirics of Agglomeration Economies”, in Duranton, G., J.V. Henderson and W.C. Strange, *Handbook of Regional and Urban Economics*, Vol. 5, Ch.5, pp. 247-348.
- Combes, P.P., G. Duranton and L. Gobillon (2016), “The Costs of Agglomeration: House and Land Prices in French Cities”, CEPR Discussion Paper, No. 9240, Originally published December 2012, updated June 2016.
- Combes, P.P., G. Duranton and H.G. Overman (2005), “Agglomeration and the adjustment of the spatial economy”, *Papers in Regional Science*, 84(3), pp. 311-349.
- Criscuolo, C. et al. (2016), “The Causal Effects of an Industrial Policy”, Revised version, *CEP Discussion Paper*, No.1113.
- De la Roca, J. and D. Puga (2017), “Learning by working in big cities”. *The Review of Economic Studies*, Vol. 84(1), pp. 106-142.
- Dehejia, R.H. and S. Wahba (1999), “Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs”, *Journal of the American Statistical Association*, Vol. 94(448), pp. 1053-1062.
- Donaldson, D. (forthcoming), “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure”, *The American Economic Review*.
- Duflo, E. (2006), “Field Experiments in Development Economics”, in Blundell, R., W. Newey and T. Persson (eds.) *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, Cambridge University Press.

- Duflo, E., R. Glennerster and M. Kremer (2007), “Using Randomization in Development Economics Research: A Toolkit”, in T.P. Schultz and J.A. Strauss (eds.) *Handbook of Development Economics*, Vol. 4, Ch. 61, pp. 3895–3962.
- Duranton, G. and D. Puga (2004), “Micro-foundations of urban agglomeration economies”, in J. V. Henderson, J.V. and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Vol. 4, Ch. 48, pp. 2063-2117.
- Duranton, G. and M.A. Turner (2012), “Urban Growth and Transportation”, *Review of Economic Studies*, Vol. 79, pp. 1407-1440.
- Einio, E. and H.G. Overman (2016) “The (displacement) effects of spatially targeted enterprise initiatives: Evidence from UK LEGI”, *CEPR Discussion paper*, DP 11112.
- European Commission (2017), *State Aid Scoreboard 2015*, available at http://ec.europa.eu/competition/state_aid/scoreboard/index_en.html (accessed 6 March 2017).
- European Commission (2014a), “You can’t improve what you can’t measure: State aid and evaluation”, *Occasional discussion paper by the Competition Directorate*, Issue 7.
- European Commission (2014b), “Investment for jobs and Growth: Promoting development and good governance in EU regions and cities”, *Sixth report on economic, social and territorial cohesion*.
- Faber, B. (2014), “Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System”, *Review of Economic Studies*, Vol. 81(3), pp. 1046-1070.
- Fryer Jr., R.G. (2016), “The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments”, *NBER Working Paper*, No. 22130.
- Garcilazo, E., J. Oliveira Martins and B. Thompson (2010), “Why policies may need to be place-based in order to be people-centred”, VoxEU. Available at <http://voxeu.org/article/why-policies-may-need-be-place-based-order-be-people-centred> (accessed 15 September 2016).
- Gibbons, S. and H.G. Overman (2012), “Mostly pointless spatial econometrics? ”. *Journal of Regional Science*, 52(2), 172-191.
- Gibbons, S., M. Nathan and H.G. Overman (2014), “Evaluating spatial policies”. *Town Planning Review*, Vol. 85(4), pp. 427-432.
- Gobillon, L. and T. Magnac (2016), “Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls”, *The Review of Economics and Statistics*, Vol. 98(3), pp. 535-551.
- Gordon, I. (1999), “Targeting a leaky bucket”, *New Economy*, Vol. 6(4), pp. 199-203.
- Greenstone, M., R. Hornbeck and E. Moretti (2010), “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings”, *Journal of Political Economy*, Vol. 118(3), pp. 536-598.
- Griffith, R., S. Lee, and J. Van Reenen (2011), “Is distance dying at last? Falling home bias in fixed-effects models of patent citations”, *Quantitative Economics*, 2(2), 211-249.

- Guiso, L., P. Sapienza, and L. Zingales (2010), “Civic capital as the missing link” *NBER Working Paper*, No. 15845.
- Hanushek, E.A. (1999), “Some Findings From an Independent Investigation of the Tennessee STAR Experiment and From Other Investigations of Class Size Effects”, *Educational Evaluation and Policy Analysis*, Vol. 21, pp. 143-163.
- Henderson, J.V., A. Storeygard and D.N. Weil (2012), “Measuring Economic Growth from Outer Space”, *American Economic Review*, 102(2), 994-1028.
- Kerr, W.R. and S.D. Kominers (2015), “Agglomerative Forces and Cluster Shapes”. *Review of Economics and Statistics*, Vol. 97(4), pp. 877-899.
- Kling J.R., J.B. Liebman and L.F. Katz (2007), ”Experimental Analysis of Neighborhood Effects”,. *Econometrica*, Vol. 75(1), pp. 83-119.
- Magrini, S. (1999), “The evolution of income disparities among the regions of the European Union“, *Regional Science and Urban Economics*, 29(2), 257-281.
- Magrini, S. (2004), “Regional (di)convergence”, in J. V. Henderson, J.V. and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Vol. 4, Ch. 62, pp. 2741-2796.
- Manski, C.F. (1993), “Identification of endogenous social effects: The reflection problem”, *The review of economic studies*, 60(3), 531-542.
- Marshall, A. (1890), *Principles of political economy*, Maxmillan, New York.
- McCann, P. and A. Rodríguez-Pose (2011), “Why and When Development Policy Should be Place-Based”, in OECD (2011), *OECD Regional Outlook 2011: Building resilient regions for stronger economies*, OECD Publishing.
- McGill, I. (2011), “Improving Regional Development Policies”, in OECD (2011), *OECD Regional Outlook 2011: Building resilient regions for stronger economies*, OECD Publishing.
- Miguel, E. et al. (2014), “Promoting Transparency in Social Science Research”, *Science*, Vol. 343(6166), pp. 30-31.
- Moretti, E. (2012), *The New Geography of Jobs*. Houghton Mifflin Harcourt.
- Monfort, P. (2008), “Convergence of EU regions: Measures and evolution”, *European Union Regional Policy Working Paper*, No. 01/2008.
- Nathan M., and H.G. Overman (2013), “Agglomeration, clusters, and industrial policy”, *Oxford Review of Economic Policy*, Vol. 29(2), pp. 383-404
- Neumark, D. and H. Simpson (2015), “Place-Based Policies”, in Duranton, G., J.V. Henderson and W.C. Strange (eds.), *Handbook of Regional and Urban Economics*, Vol. 5, Ch. 18, pp. 1197-1288.
- OECD (2016a), *OECD Regional Outlook 2016: Productive Regions for Inclusive Societies*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264260245-en>.

- OECD (2016b), *The Productivity-Inclusiveness Nexus*, Background document for the 2016 Meeting of the OECD Council at Ministerial Level, Paris, 1-2 June 2016, available at: <http://www.oecd.org/mcm/documents/The-productivity-inclusiveness-nexus.pdf>
- OECD (2015a), *OECD Regulatory Policy Outlook 2015*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264238770-en>.
- OECD (2015b), *The Metropolitan Century: Understanding Urbanisation and Its Consequences*. OECD Publishing, Paris.
- OECD (2015c), *Government at a Glance 2015*, OECD Publishing, Paris. http://dx.doi.org/10.1787/gov_glance-2015-en
- OECD (2015d), *The Innovation Imperative: Contributing to Productivity, Growth and Well-Being*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264239814-en>
- OECD (2015e), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264229358-en>
- OECD (2013), *OECD Regions at a Glance 2013*, OECD Publishing, Paris. http://dx.doi.org/10.1787/reg_glance-2013-en
- OECD (2012), *Redefining “Urban”: A New Way to Measure Metropolitan Areas*, OECD Publishing, Paris.
- OECD (2011), *OECD Regional Outlook 2011: Building resilient regions for stronger economies*, OECD Publishing.
- Oldsman, E. (2014), “Making evaluations count: Toward more informed policy”, in: OECD/The World Bank (2014), *Making Innovation Policy Work: Learning from Experimentation*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264185739-11-en>
- Overman, H.G. (2004), “Can we learn anything from economic geography proper?”, *Journal of Economic Geography* 4(5), pp. 501-516.
- Puga, D. (2010), “The magnitude and causes of agglomeration economies”, *Journal of Regional Science*, Vol. 50(1), pp. 203-219.
- Puga, D. (2002), “European regional policies in light of recent location theories”, *Journal of Economic Geography*, Vol. 2(4), pp. 373-406.
- Rappaport, J. (2007), “Moving to nice weather”, *Regional Science and Urban Economics*, Vol. 37(3), pp. 375-398.
- Rathelot, R. and I. Marinescu (2014), “Mismatch Unemployment and the Geography of Job Search”, mimeo.
- Redding, S. and M.A. Turner (2015), “Transportation Costs and the Spatial Organization of Economic Activity”, in Duranton, G., J.V. Henderson and W.C. Strange, *Handbook of Regional and Urban Economics*, Vol. 5, Ch.20, pp. 1339-1398.

- Rodrik, D. (2008), *Normalizing industrial policy*. International Bank for Reconstruction and Development / The World Bank.
- Rubin, D.B. (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies”, *Journal of Educational Psychology*, Vol. 66(5), pp. 688–701.
- Rubin, D.B. (1986), “Statistics and Causal inference: comment: which ifs have causal answers”. *Journal of the American Statistical Association*, Vol. 81, pp. 961–962.
- Schumann, A. (2016), “Using Outcome Indicators to Improve Policies: Methods, Design Strategies and Implementation”, *OECD Regional Development Working Papers*, No. 2016/02, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jm5cgr8j532-en>.
- Teixeira, A.C. (2006), “Transport policies in light of the new economic geography: The Portuguese experience”, *Regional Science and Urban Economics*, Vol. 36(4), pp. 450–466.
- Tobler, W. (1970), “A computer movie simulating urban growth in the Detroit region”. *Economic Geography*, Vol. 46(2), pp. 234–240.
- Warwick, K. and A. Nolan (2014), “Evaluation of Industrial Policy: Methodological Issues and Policy Lessons”, *OECD Science, Technology and Industry Policy Papers*, No. 16, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5jz181jh0j5k-en>