FISCAL FEDERALISM IN THE 21ST CENTURY

David R. Agrawal
Jan K. Brueckner
Marius Brülhart

Abstract

Fiscal federalism concerns the division of policy responsibilities among different levels of government. Many current economic and policy developments, such as globalization, environmental crises and rising inequality, may not appear to be favorable to fiscal federalism, yet countries are further decentralizing their fiscal systems. We summarize the efficiency and equity aspects fiscal decentralization, fiscal competition, fiscal externalities, and intergovernmental grants. The survey introduces readers to theoretical reasons for/against a federalist structure. We discuss how federalism relates to classic problems in economics: externalities, inequality, spillovers, information, and aspects of political economy. Our survey integrates both theory and empirics, while also focusing on the variety of federal systems in different countries, both developing and developed. We conclude by discussing how fiscal federalism is being shaped by economic, technological, and environmental changes, while discussing the effects of globalization, polarization, and global crises on the future of federal systems.

Keywords: fiscal federalism, intergovernmental relations, grants, fiscal competition, fiscal externalities, decentralization

---

David R. Agrawal: University of Kentucky; Martin School of Public Policy and Department of Economics; 433 Patterson Office Tower; Lexington, KY, 40506-0027 USA; +1-859-257-8608; dragrawal@uky.edu; orcid: 0000-0001-8480-1849
Jan K. Brueckner: University of California, Irvine; Department of Economics; 3151 Social Science Plaza; Irvine, CA 92697 USA; +1-949-824-0083; jkbrueck@uci.edu; orcid: 0000-0001-9580-1924
Marius Brülhart: University of Lausanne; Department of Economics, HEC Lausanne; CH – 1015 Lausanne Switzerland; +41-21-692-3471; Marius.Brulhart@unil.ch; orcid: 0000-0001-5483-0219

Agrawal and Brueckner are Fellows of CESifo and Brülhart is a fellow of CEPR.

Disclosure: The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgments: We thank Tim Besley, Robin Boadway, Jeffrey Clemens, William Hoyt, Julia Payson, David Wildasin, and an anonymous reviewer for comments and discussions on earlier drafts of the paper. Chinmay Korgaonkar provided excellent research assistance. DOI: https://doi.org/10.1146/annurev-economics-081623-020713
1. Introduction

In a fiscal federation, tax and spending decisions are not confined to the national government but are shared with subnational jurisdictions. How to allocate fiscal decision-making powers across central, state, and local governments is a matter of unending scientific and political debate. Over recent decades, the mean sub-federal share of total tax revenue has been gradually increasing within both OECD and non-OECD countries.¹ But, substantial cross-sectional variation in the extent of decentralization and the number of subnational tiers remains. Federalism is a continuum that varies both by the extent of sub-federal fiscal autonomy and by the number of jurisdictional tiers, ranging from multi-tiered systems like the U.S. and Canada with governance at the central, state, county, and municipal level to countries with mainly municipal decentralization.

Many current policy debates concern issues of fiscal federalism. What level of government should oversee disaster policy, such as hurricane-damage mitigation or control of an infectious disease?

¹ According to the IMF’s Government Finance Statistics, Canada was the least fiscally centralized country in 2019, with a sub-federal tax share of 55% (up from 51% in 2000), followed by Switzerland (53%, up from 51%) and the U.S. (47%, up from 40%). Across countries in the database, the mean sub-federal tax share in 2016-2019 was 20% among OECD countries and 13% among non-OECD countries, confirming that developing nations tend to be fiscally more centralized (Gadenne and Shingal, 2014). In this paper, we consider different layers of sub-national government interchangeably. For a survey that focuses specifically on decentralization to the lowest, i.e. local, layer, see Brülhart et al. (2015).
disease? How progressive should state income taxes be, and is inequality best solved at the local or federal level? What role do redistributive grants play?

But many economic, policy and technological developments—globalization, rising inequality, and global challenges such as climate change—may not appear to be favorable to fiscal federalism. Although globalization has increased the mobility of goods, capital and people, leading to calls for supranational tax reforms and coordination, these forces have also led to backlashes resulting in local protectionism. Global challenges such as climate change or infectious-disease control similarly are seen as requiring national and supranational coordination, but information asymmetries combined with political realities create reasons for local intervention. Finally, growing economic inequality in many countries may be best addressed by national governments in the presence of mobility, but redistributive policy may also be viewed as a local public good.

There are also forces that favor decentralization. In the face of political polarization and gridlock, voters have sought to bring governance closer to home (Boadway and Shah 2009). Examples from the U.S. include recent state proposals on immigration policy or wealth taxation following failed federal alternatives, or court decisions that granted states more autonomy over social policy, with the federal legislature reluctant to establish limitations. Multi-layered governance allows lower-level governments to foster citizen engagement, better set policies to match voter preferences, and exploit other elements of local information, but it may create interjurisdictional externalities and spillovers. Federalism has also come to be viewed as a tool for improved governance and reduced corruption in developing countries.

In this survey, we summarize recent research on classic questions in federalism, including the benefits and costs of decentralization, the allocation government tasks to central or local governments, fiscal externalities, the role of redistributive policy, and the political economy of
federalism. We also explore how fiscal federalism is being shaped by economic, technological, and environmental changes. Obviously, we must be selective in the topics we cover.

The literature on fiscal federalism was last comprehensively surveyed in Oates (1999). Our review will make clear that many classic models of federalism remain highly relevant, although recent work has placed some qualifications on these models, especially in the context of low-income countries. Further, federalism has been reshaped by many more-recent events, including the pandemic, which spurred both calls for and resistance to centralized policies while at the same time leading to increased federal support for individuals and sub-national governments. We will highlight how recent trends have refined our understanding of federalism. In doing so, our emphasis is on understanding the efficiency and equity tradeoffs of fiscal federalism that are critical for the design of governance structures and for the evaluation of public policies.

2. Heterogeneous public-good demands and fiscal decentralization

The benefits of fiscal decentralization in economic models arise because of a standard assumption used in almost all analysis of public goods: consumption of a public good must be the same for all residents of the jurisdiction providing the good, although it can vary across jurisdictions. Because educational quality can vary across schools or the effectiveness of police protection can vary across a city’s neighborhoods, the assumption of uniform consumption within a jurisdiction is not entirely realistic. But it serves as a useful approximation in the analysis of fiscal federalism.

The consumption-uniformity condition creates a stark contrast between centralized and decentralized provision of public goods. Under centralized provision, the consumption of a public good (denoted z) must be the same for all residents of the national jurisdiction. Decentralized provision, however, allows consumption levels to vary across the country’s subnational
jurisdictions, which can provide different levels of the public good. With population sorting, these jurisdictions may then respond to heterogeneous demands for public goods, with some jurisdictions providing a high $z$ level to high demanders and others providing a low $z$ to low demanders.

This insight was the basis for the famous paper of Tiebout (1956), written in response to Samuelson (1954), who stressed the inefficient compromise inherent in providing a common level of $z$ to consumers with different preferences. The Tiebout model’s idea that subnational provision of $z$ allowed the public sector to respond to demand heterogeneity was elaborated in Oates (1972), and its limitations discussed by Boadway and Tremblay (2012).

To generate the social optimum under decentralized provision of a public good, a social planner would organize consumers into demand-homogeneous subnational jurisdictions of optimal size. This size accounts for returns to scale in production of $z$ as well as the good’s congestibility. The latter property captures the extent to which an increase in community population degrades consumption of the good, holding resources devoted to producing $z$ constant. If a particular demand group has a large population, many homogenous optimal-size communities may be required to accommodate it.

While this planning solution is well understood (Berglas and Pines 1981), the institutional structure required to achieve it as a decentralized equilibrium has been a matter of debate. For example, Brueckner (1979) showed that, when the level of $z$ is decided by majority voting under a head-tax regime and consumers are free to move between communities, inefficient equilibria may exist, containing mixed (rather than homogeneous) communities of non-optimal sizes. More recently, Calabrese, Epple and Romano (2012) demonstrated a similar conclusion in a detailed computational model. Using a framework where the public good is financed by a property tax and
where discrete consumer types are replaced by a continuum with different incomes and preferences, the analysis demonstrated that majority voting along with free mobility again may generate inefficient equilibria.

Hence, “voting with one’s feet,” as captured by the free mobility assumption, is not enough to generate the efficient decentralized equilibria envisioned by Tiebout (1956). Subsequent researchers built on elements of Tiebout’s argument by recognizing that another class of optimizing agents is needed to guide an economy toward efficient equilibria: a class of competitive community developers. These developers charge a community entry fee that finances provision of $z$ while choosing the level of that good along with community population to maximize profit. This approach was first developed by Berglas (1976, 1981) using a “utility taking” approach that, in effect, required developers to know consumer preferences. Multiple taste groups made this approach untenable, since developers would then have the unrealistic ability to distinguish among groups and charge them different fees. A superior price-taking approach, however, was developed by Scotchmer and Wooders (1987). Under this approach, developers and consumers both face a parametric price function that expresses the community entry fee as a function of $z$. An equilibrium price function, which “clears” the market for public goods (with profit-maximizing communities yielding the $z$ levels demanded by consumers), then generates an efficient equilibrium, in which consumers are sorted into homogeneous, zero-profit communities with optimal $z$ levels, without the violation of consumer anonymity inherent in the utility-taking approach (see also Scotchmer 2002; Wooders 1978).

While some parts of rapidly growing US cities (and perhaps cities in China) have been developed at a large scale in a manner roughly matching the community-developer model, the model is generally unrealistic. But the framework can be viewed as providing a benchmark
institutional structure that generates efficient equilibria in the presence of public goods. In the same way, the (generally unrealistic) model of competitive markets is a benchmark that yields efficient equilibria in basic economic theory. Note, while the absence of a downside makes decentralization preferred to centralized provision in the discussion so far, additional elements introduced below may overturn this conclusion.

3. Empirical tests for voting with one’s feet

The homogenization of communities via voting with one’s feet has powerful intuitive appeal, which has led to several empirical tests. Rhode and Strumpf (2003) argue that voting with one’s feet should lead to an increase over time in intercommunity differences in taxes and per capita spending, and in demand-related demographic characteristics, as consumers self-segregate. Their analysis finds exactly the opposite conclusion, showing that communities have become more, not less, similar to one another over time. The authors note that, because the theory ignores forces that favor mixed communities, such as labor complementarity in production (which may require collaboration of workers whose public-good demands differ), the prediction of demand homogenization may be overstated, perhaps helping to explain their findings.²

Eberts and Gronberg (1981) focus on income variation within school districts, arguing that, as the number of districts in a metro area increases, allowing greater scope for homogenization, income variation within districts should decrease. Using an entropy measure that decomposes a metro area’s income variation into within-district and across-district components, a regression of the within-district component on the number of districts in the metro area (treated as endogenous)

² Public-good models with complementarities across demand groups show the optimality of mixing. See, for example, Brueckner (1994) and Conley and Wooders (1997).
and other covariates yields a negative coefficient. This finding confirms a tendency toward the homogenization of jurisdictions as their number increases.

Banzhaf and Walsh (2008) carry out a more literal test for voting with one’s feet by providing evidence of consumer migration in response to changing exposure to pollution, not a traditional public good. They show evidence of migration into areas that experience reductions in the level of toxic industrial pollutants, showing that consumers do vote with their feet when environmental factors change. Finally, Epple and Sieg (1999) execute an indirect test for voting with one’s feet using a structural empirical model. They ask whether interjurisdictional conditions for consumer locational equilibrium are satisfied, reaching an affirmative answer.

4. Uninternalized public-good spillovers

The models of decentralization considered so far omit the possibility of public-good spillovers across jurisdictions. When spillovers exist, residents benefit from public goods in adjacent communities as well as those provided by their own community. As an example, residents might travel to an adjacent community to visit its parks. As pointed out by Oates (1972), when the level of \( z \) is chosen in a decentralized structure accounting for only local benefits and costs, the good will be underprovided when spillovers exist, requiring intervention by the central government. With \( z \) chosen by majority vote (or perhaps by community developers), underprovision can be corrected by a system of matching grants from the central government, which reduces the effective cost of \( z \) (see also Section 10.3). With the cost of the public good lowered, decision makers then expand the level of provision. The matching-grant rate reflects the extent of spillovers, being larger with high spillovers, thus encouraging an appropriate increase in provision, as discussed by Oates (1972).
In contrast to the previous conclusions, the need for central intervention in the presence of spillovers vanishes in the model analyzed by Wellisch (1993). He builds on work by Myers (1990), who portrays two communities as maximizing their identical residents’ utility in strategic fashion, accounting for intercommunity migration induced by changes in public-good provision. Myers shows that communities have an incentive to make voluntary transfers to one another, which in effect forestall unwanted intercommunity migration and lead to an efficient Nash equilibrium (in contrast to the inefficiency shown by Flatters, Henderson and Mieszkowski 1974). Generalizing this framework, Wellisch (1993) demonstrates that the Nash equilibrium with transfers yields efficient public-good levels even in the presence of spillovers, obviating the need for central intervention. With Myers’ transfers forestalling migration “spillovers,” it is natural that transfers also internalize another type of spillover, one involving public goods.

Ogawa and Wildasin (2009) demonstrate a similar result in a tax-competition model where capital, the mobile taxed factor, generates a harmful byproduct, such as pollution, that spills across community borders. When the spillover is imperfect, the community benefits from raising its tax, which drives capital away and thus reduces (via the imperfect spillover) consumption of the harmful byproduct. Ogawa and Wildasin show that this effect offsets the usual incentive for a tax reduction under tax competition, leading to an efficient outcome, with no need for central intervention. While the arguments of Wellisch (1993) and Ogawa and Wildasin (2009) are conceptually noteworthy, they perhaps represent idiosyncratic exceptions to the general view that spillovers lead to inefficiency. But at the same time, they provide insights for when federalist structures can be useful in addressing global challenges discussed in the introduction.

Trans-border pollution externalities can lead to spillovers in environmental policies across jurisdictions, with tighter environmental regulation by a jurisdiction’s neighbors reducing local
pollution. River pollution is an important source of such externalities, as documented and analyzed by Sigman (2002) and Lipscomb and Mobarak (2017). Researchers have sought evidence of environmental-policy spillovers by looking for evidence of strategic interaction across jurisdictions in the choice of such policies, as in Fredriksson and Millimet (2002). Studies that look for spillover-generated strategic interaction in spending on other public goods include Case, Rosen and Hines (1993) and Solé-Ollé (2006).

5. Deciding on decentralization in the presence of both benefits and costs

If efficiencies from spillovers are not remedied via matching grants, is fiscal decentralization desirable? Besley and Coate (2003) provide a transparent answer in a specialized model with two communities, each containing a continuum of consumer valuations for $z$, whose means differ across communities. They show that if spillovers are weak, then the gains from tailoring the public good to suit community preferences exceed the losses from underprovision due to spillovers, making decentralization preferable. On the other hand, centralized provision of the public good, which takes spillovers into account, is superior when spillovers are strong.

The same yes-or-no question can be asked when decentralization involves costs other than uncorrected spillovers. One such case arises when the public good is financed by distortionary taxes on a mobile factor, as under horizontal tax competition. As mentioned above, for example, jurisdictions attempt to avoid a loss of mobile capital (which confers a positive horizontal fiscal externality on neighbors) by keeping the capital tax low, an outcome that typically leads to inefficient underprovision of the public good (Agrawal, Hoyt and Wilson, 2022). This distortion is mitigated under centralized provision because capital flight across subnational jurisdictions does not occur when a tax increase occurs at the national level.
To ask whether decentralization is desirable in this setting, Brueckner (2004) uses a parameterized model with two demand groups that are exogenously separated into two homogeneous communities. Examples show that decentralization is desirable when the preferences of the two groups are sufficiently different (enhancing the gain from a non-uniform $z$) or when capital flight in response to a higher tax rate is sufficiently weak (mitigating the loss from tax competition). This limited flight requires a high curvature of the private-good production function.

Instead of modeling decentralization as a discrete choice, Janeba and Wilson (2011) treat the decentralization decision in the presence of tax competition as continuous. They do so by assuming a continuum of public goods, with goods above an endogenous critical point in the continuum provided by the central government and goods below the critical point provided by subnational governments. Some degree of decentralization is desirable when the optimal critical point is positive. Since they assume common preferences, decentralization entails the losses from tax competition but no benefits. Centralization also involves an inefficiency loss due to the capture of central decision-making by a subset of jurisdictions, who favor themselves in public spending. In answering the decentralization question, the analysis then weighs these two sources of inefficiency.

6. Decentralization of income redistribution

Although Musgrave (1959) believed that income redistribution should be a national responsibility, Pauly (1973) argued that, because the altruism underlying income redistribution may be spatially limited, subnational redistribution may be superior. In other words, if the altruism toward the poor among well-off residents in a state like California extends only to the poor residents of that state, whose circumstances may be visible on a regular basis, then redistribution
to the California poor should be done within the state, not at the national level. Such decentralized redistribution allows the extent of transfers to the poor to vary across states, recognizing possible differences in the strength of the well-off population’s altruism.

However, when redistribution is decentralized, tax-induced migration may arise as the rich move to less-progressive states and the poor relocate to states with more-generous transfers. Such migration can be viewed as the mirror image of capital flight in tax-competition models, and it has a similar effect on policy decisions. While the threat of capital flight is expected to keep taxes low, leading to underprovision of public goods, the migration threat may also keep transfers low as states seek to avoid the burden of a larger poor population and as the mobility of the wealthy raises the costs of redistributive taxation. Migration then may prevent a state from setting transfers at level consistent with its degree of altruism.

US welfare reform in the 1990s, which constrained the poor’s eligibility for transfers, made welfare migration less of a policy issue. But many empirical studies from the pre-reform era attempted to measure the extent of welfare migration, with most studies finding affirmative evidence (Brueckner 2000). However, even as welfare migration has declined, growing inequality combined with progressive state and local taxes on income and wealth has spurred a new literature on tax competition for the elite, given that migration of the wealthy is an important policy issue (Kleven et al. 2020).

Theoretically, the effect of migration on transfers was analyzed by Wildasin (1991), who showed that the underprovision of transfers can be corrected by matching grants from the central government. These grants reduce the cost of welfare to the states, encouraging more redistribution and making it possible to efficiently implement Pauly’s (1973) amendment to Musgrave’s (1959) redistribution strategy.
The matching-grant remedy is very similar to the one suggested as a correction for public-good spillovers, although the underlying mechanism creating the distortion is different. While it is unlikely that the previous logic was the reason, the pre-reform US welfare system used a system of federal matching grants, with matching rates differing across states. Unfortunately, the system was converted to one of block (lump-sum) grants following the reform, a change that was undesirable from the perspective of the foregoing analysis.

Migration in a decentralized system with income transfers can be a source of strategic interaction among jurisdictions, with each looking at its neighbors’ transfer or income tax levels in making its own choice. Tests for such strategic interaction were carried out by Figlio, Kolpin and Reid (1999), Saavedra (2000), Dahlberg and Edmark (2008), and Brueckner (2023), with each study finding evidence of the phenomenon.

7. Alternative portrayals of fiscal decentralization

7.1 Partial fiscal decentralization

The regimes of fiscal decentralization discussed so far assume that subnational governments are fully autonomous in their taxing and spending decisions. In reality, however, decentralization in many countries is more limited, with subnational decisions partly under the control of the central government. Such arrangements may be due to constitutional restrictions or limited tax capacity at the subnational level, with jurisdictions not having productive taxes at their disposal. Alternatively, central control may involve a desire to partly dictate spending that is deemed socially desirable. Such an arrangement, where public-sector decision making is divided between central and subnational governments, has been called partial fiscal decentralization by Brueckner (2009).
Suppose that subnational jurisdictions have no revenue sources whatsoever, relying entirely on central-government grants for funds, and that they have identical populations and receive equal grants. Then with a single public good and identical subnational production functions for \( z \), this partial decentralization arrangement is identical to centralized provision since the common grant size automatically dictates a common subnational \( z \) level.

One way of breaking this equivalence, allowing a scope for partial decentralization, is to assume that the cost of \( z \) depends on the effort expended by local bureaucrats, so that a given grant can translate into different \( z \) levels depending on the amount of effort exerted. Using such a model, Brueckner (2009) shows, among other things, that when chosen optimally, a uniform central grant financed by a national head tax leads not to a uniform \( z \) but to a narrower range of \( z \) than if the economy’s homogeneous communities were to rely on their own individualized taxes. Partial decentralization in this case constrains the local public sector’s response to the diversity of demands for \( z \). Under a different portrayal of partial decentralization proposed by Borge, Brueckner and Rattsø (2014), the central government again supports subnational spending via grants, but funding is for a collection of multiple public goods rather than a single good. With partial decentralization, the central government relaxes mandates that dictate a fixed division of grant spending by localities across the different public goods, allowing jurisdictions to adjust the spending mix according to local preferences.

Using a model like that of Janeba and Wilson (2011), Hatfield and Padró i Miguel (2012) study the problem of partial decentralization as the choice of a critical point in a continuum of public goods. In contrast to Janeba and Wilson’s normative approach, the authors provide a political-economy analysis, with the critical point chosen by the economy’s median voter. Ownership and investment of mobile capital, however, adds an analytical layer, although the
timing assumptions in the model eliminate the tax-competition distortion present in Janeba and Wilson’s model.

In empirical work on partial decentralization, Faguet (2004) finds that, after a Bolivian reform increased the size of grants and gave localities more control over the investment projects they financed, investment levels changed in ways that better reflected local demand characteristics. Similarly, studying a Norwegian reform that relaxed mandates on the division of fixed grant funds across different public goods (as described above), Borge, Brueckner and Rattso (2014) found that the levels of spending on the goods became responsive to local demand characteristics in a fashion that was not observed before the reform.

7.2 Central provision of non-uniform public-good levels

After providing an illustration of how the decentralization decision depends on the extent of spillovers in a standard model, Besley and Coate (2003) go on to relax the key assumption that underlies almost all analysis of fiscal federalism: uniformity of z levels under centralized provision. One of their goals is to assess how the role of spillovers in the decentralization decision is affected by this change, but exploration of the impact of this new assumption has broader value. Lockwood (2002) provides a similar analysis.

The outcome of centralized provision, where nonuniform z levels can be chosen, depends on how the central legislature aggregates the preferences of the elected representatives from the two subnational jurisdictions. Under the noncooperative approach, a representative chosen at random picks the z levels in both of the two jurisdictions. Under the cooperative approach, the levels are chosen to maximize joint surplus of the two elected representatives.

The preferences of the two elected representatives are set by the median voter in each of the jurisdictions, following the citizen-candidate approach of Besley and Coate (1997). Under the
noncooperative approach, the representative sent to the central legislature has the median voter’s own preferences, while under the cooperative approach, the representative’s preferences diverge from those of the median voter.

Under the noncooperative approach, the randomly chosen representative sets the \( z \) levels in both jurisdictions to reflect her own preferences, adjusted for spillovers. When spillovers are perfect, the representative sets equal \( z \) levels in the two jurisdictions, but when spillovers are imperfect, \( z \) is set higher in the representative’s own jurisdiction than in the other jurisdiction. If spillovers are entirely absent, however, the other jurisdiction gets a zero \( z \) level, since that jurisdiction’s level then provides no benefit to the decision maker. The cooperative approach, by contrast, reflects a spirit of compromise in each spillover case.

Turning to the decentralization question, the decision under the noncooperative approach, which focuses on expected surplus, turns out to use the same criterion as in Besley and Coate’s version of the standard model: decentralize when the extent of spillovers lies below a critical level and centralize otherwise. The analysis of decentralization in the noncooperative model is more complex, but it yields a decision rule that diverges only slightly from the noncooperative one.

8 Multi-level governance

Under a federal system, different tiers of government (federal, state, county, local) may share the same revenue sources and expenditure/regulatory tasks. In the case of revenues, a given tax base may be the common property of multiple levels of government. In the U.S., for example, both the federal government and state governments levy their own taxes on income, gasoline, and cigarettes. The issues arising from the co-occupancy of tax bases also arise with respect to provision of public services.
8.1 Theory of vertical externalities and vertical fiscal competition

Section 4 briefly discussed horizontal tax competition, the process by which jurisdictions at the same level of the federal system compete for mobile tax bases. The horizontal externalities stemming from this mobility of the tax base generally imply that a tax cut in one jurisdiction harms the tax base of other jurisdictions. As Agrawal, Hoyt and Wilson (2022) previously surveyed horizontal fiscal externalities, our focus is instead on interdependencies between different tiers of government. Vertical tax setting is the process by which governments set tax rates taking account of taxes imposed by other levels of government. Vertical fiscal externalities between these different levels likely have an effect opposite to that of horizontal externalities: a federal tax increase reduces the tax base of lower-level governments via changes in demand for labor, goods, or factors. Thus, while horizontal externalities often imply taxes are too low, vertical tax externalities may imply that they are too high (Keen and Kotsogiannis 2003).

Consider an example of a tax on consumption levied by both the federal government and the states. Much of the prior literature assumes that each government, in choosing a tax rate, considers only the effect on its own tax revenues (Boadway, Marchand, and Vigneault 1998). Then, following Dahlby (1996), a federal government considering raising its tax rate will set the marginal benefits from the added revenue equal to the marginal cost of public funds (MCPF)–or the marginal loss due to increased taxes. But, since an increase in the federal government’s tax rate reduces the quantity of goods demanded, shrinking the tax base of the state governments, the social MCPF differs from the MCPF perceived by the federal government. The federal government then overtaxes consumption.

This assumption that governments care only about their own revenue can be relaxed. If so, the federal government considers the impact on state revenues, but a state only partially considers
its effects on federal revenues (Hoyt 2001). In such models, states are only concerned with their impact on federal revenues insofar as those federal revenues fund services provided to residents of their state and states are not concerned with the impact on federal services provided to other states. As lower-level governments become more fragmented, the externality any one jurisdiction imposes on the federal government is small, but they also care less about their impact on the federal government.

Vertical fiscal externalities also arise in the provision of public expenditures by state and federal governments. For example, if public spending at the federal level (say on education subsidies) has economic productivity effects at the state level, the resulting increase in wages will raise the state income tax base (Wrede 2000; Dahlby and Wilson 2003). As the above analysis makes clear, the concept of vertical externalities shares many similarities to the common pool problem in private markets (e.g., fisheries).

In response to vertical externalities imposed by the federal government, lower-level governments may strategically adjust their policies. If taxes are set strategically, does a state government raise or lower its tax rate in response to a federal tax-rate increase? To build intuition, we follow the simple case where governments maximize their own tax revenue by setting taxes (Keen 1998). Then, using the example of an excise tax following Keen (1998), state governments maximize $tx(q)$ where $t$ is the state tax rate and $x(q)$ is demand as a function of the after-tax price $q = p + t + T$, inclusive of the federal tax rate $T$, and where $p$ is a constant producer price. Governments then set taxes following the usual Ramsey rule: $t/q = 1/e(q)$ where $e(q) > 0$ is the elasticity of demand. If the elasticity of demand is constant, then we can unambiguously conclude that federal and state taxes are strategic complements, with $dt/dT > 0$. But, if demand is linear, then $x'(q)$ is constant and an increase in $T$ implies state taxes must fall for the Ramsey
rule to continue to hold, yielding $dt/dT < 0$. Put differently, while a tax increase at one level usually harms the tax base of other levels of government, the strategic reactions between a federal and sub-federal government may result in best response functions that slope up or down, an ambiguity also seen under horizontal tax competition. The relative response of one government to another will depend critically on the elasticity of demand. However, if governments maximize welfare, it is more likely the reaction function slope is positive.

What level of government should tax particular bases or provide particular services? Summarizing this “assignment problem”, Dahlby (2001) argues that lower-level governments should tax bases with less interjurisdictional mobility and bases that are less volatile, while higher-level governments should tax bases that are distributed unequally across jurisdictions, that are useful for stabilization policy, and that have progressive equity elements. However, Dahlby does not provide any formal guidelines about how the assignment problem is influenced by the issue of co-occupancy nor whether co-occupancy is even advisable.³

The early literature suggested separating tax bases between different tiers (Flowers 1988), decreasing the number of lower-level governments (Keen and Kotsogiannis 2004), or providing intergovernmental grants to correct for the vertical externalities (Boadway and Keen 1996). But

³ The assignment problem is at the heart of many debates about whether the European Union should become more centralized at the supra-national level or should leave policies decentralized at the level of its Member states. The principle of “subsidiarity” is a compromise where member states maintain their autonomy unless EU policy is explicitly considered necessary. This is mainly a legal and political concept, though, and may not imply a clear assignment of policies based on economic criteria (for a survey, see Gelauff, Grilo and Lejour 2008).
there may be reasons for co-occupancy, such as in Haufler and Lulfesmann (2015), where co-occupancy can help correct horizontal externalities. Indeed, Hoyt (2017) shows that with multiple tax instruments, the elimination of co-occupancy of tax bases will not eliminate vertical fiscal externalities when the separate tax bases are interrelated. Starkly, in the presence of interdependent tax bases, if a tax increase on one base increases a related base (when they are gross substitutes), co-occupancy may be optimal.

8.2 Empirical evidence on vertical tax interactions

A recent literature tests for the existence of vertical tax interactions. Besley and Rosen (1998) find that when the federal government increases excise taxes, the states respond by increasing theirs as well, while Boadway and Hayashi (2001) find the opposite in Canada. Devereux, Lockwood and Redoano (2007) expand on this analysis by controlling for horizontal interactions, showing that failing to account for both types of tax competition may result in biased estimates of the reaction functions, and Goodspeed (2002) provides evidence that both forms of competition can interact. However, as there is only a single federal government that changes its tax rate periodically, this literature faces challenges because identification comes solely from the time series. Thus, vertical tax reactions can be hard to separate from aggregate shocks that may induce a correlation between state and federal taxes.

A more recent empirical literature on the topic (Agrawal 2015; Agrawal 2016) has resorted to exploiting multiple “federal” tax changes by re-defining the “federal” government as counties (or states) and the lower-level governments as towns within them. Agrawal (2016) uses a border discontinuity design to show that towns on opposite sides of the state border will set different local tax rates depending on the relevant state tax rates: local taxes on the low-state-tax side of the border are higher than local taxes just over the border on the high-state-tax side. This pattern points to a
negatively sloped reaction function, but the literature also finds positive or null effects at the local level. The elasticity of the tax base is a key determinant of the slope’s sign, and this elasticity and its curvature may vary dramatically across various taxes, institutions, and levels of government. Since the literature thus shows that strategic competition exists, we next discuss the implications for whether horizontal or vertical externalities dominate.

Brülhart and Jametti (2006) show theoretically that sub-federal tax rates decrease in the number of lower-level jurisdictions when horizontal externalities dominate, but that tax rates increase when vertical externalities dominate. Intuitively, similarly to Hoyt (1991), the first result arises because many small jurisdictions will compete more intensely for capital than large jurisdictions. The second result arises because smaller jurisdictions recoup a smaller share of the federal tax revenue and thus place less weight on the federal government.

Taking this model to the data, Brülhart and Jametti (2006) focus on municipalities that determine policies by direct democracy. They regress municipal tax rates on the relative size of the municipality, finding robustly positive coefficients. Thus, according to the model, vertical externalities then dominate horizontal externalities. Brülhart and Jametti (2019) confirm this result, but they find that the relationship between fragmentation and taxes becomes negative for municipalities with more delegated (elected) fiscal responsibility. Thus, in less democratic places, taxes fall as fragmentation increases. Since governments that have less-direct democracy are arguably less welfare-focused and more Leviathan in nature, horizontal tax competition may beneficially tame the tax burden imposed by these governments.

8.3 Diagonal fiscal externalities

Fiscal federations often have more than two tiers, yielding a “federation of federations.” This pattern introduces a new type of fiscal externality (Agrawal 2015). When one state raises its
tax rate, it imposes an externality not only on neighboring states, but also on the counties in those neighboring states, generating a “diagonal” externality. These externalities differ from the vertical ones discussed above because they are between governments that do not co-occupy the same base, and they also differ from horizontal externalities because they are between different tiers of government.

With respect to tax competition, counties may respond to neighboring state taxes in a manner different from neighboring county taxes, perhaps because one linkage is more salient or because some counties may be far away from the neighboring state. The strategic response may also differ if yardstick competition is at work, given that a county may use other counties as a benchmark but may not compare themselves with states. Interestingly, while Agrawal (2015) finds that the tax rates of lower- and higher-level governments co-occupying the same base are negatively related, the diagonal relationship between own-locality and neighboring higher-tier taxes is positive, as in horizontal competition. But this diagonal tax competition decays with distance to the “federation” border. Revelli, Tsai, and Zotti (2022), using the example of Italian income taxes, find that horizontal and vertical effects dominate.

9 Federal deductibility of taxes

Individuals may be allowed to deduct the state and local taxes paid on federal income tax returns. For example, in the U.S., local property taxes and state income or sales taxes are deductible from federal income taxes for individuals who itemize (the SALT deduction).

4 Parchet (2019) argues that higher-tier neighboring jurisdictions influence tax competition by affecting the localities within their own state, which influences localities in other states via horizontal competition.
Deductibility makes it more likely that states increase their taxes when the federal government raises its tax. Thus, state taxes are more likely to positively covary with federal taxes in the presence of deductibility. Esteller-More and Sole-Olle (2001) provide an empirical test of this notion by exploiting tax variation across states depending on whether the state allows federal taxes to be deducted or not on state returns. Metcalf (2011) finds deductibility increases reliance on deductible taxes and subnational spending out of own-source revenue.

The federal deductibility of state and local taxes also has important implications for horizontal externalities and the progressivity of state income taxes. Mobility of high-income taxpayers across states expands the tax base of receiving states but threatens to limit the progressivity of state income taxes, as noted in Section 6. Critically, a uniform federal subsidy will not correct this externality, because the size of the externality depends on the degree of progressivity. But the federal government can internalize the externality with a subsidy that grows with the state tax payments made by high-income residents (Cullen and Gordon 2008).

One might imagine that completely federalizing the income tax might eliminate the incentives to migrate across cities or states. But because the federal income tax is based on nominal (rather than real) income, workers with the same real incomes would pay higher taxes in high cost-of-living areas without receiving additional benefits. Albouy (2009) shows that, although wages and prices adjust across cities to compensate workers for federal tax differences, the geographic distribution of the population is inefficient. But Albouy also shows that allowing for deductions for expenditures on non-tradeable goods effectively indexes the federal income tax to the local costs, reducing these inefficiencies.
10 Fiscal equalization

Federal systems involve redistribution of public funds across sub-federal jurisdictions, and such redistribution takes a myriad of forms. One can distinguish between explicit and implicit equalization. Implicit equalization policies include regional infrastructure spending, place-based policies, and redistribution occurring through federal tax systems. In this section, we concentrate on explicit equalization, which includes schemes whose principal stated aim is to reduce disparities in public-sector resources (funds to provide ...) across sub-federal jurisdictions, and which do so by applying a given formula. Equalization can also occur as transfers facilitated through supra-national institutions, as under EU cohesion policy that is aimed at reducing regional inequalities through redistributive “place-based” transfers (for a survey, see, von Ehrlich and Overman 2020).

10.1 Tax-base equalization

Tax-base equalization, often also referred to as “fiscal capacity equalization”, has become a common system in mature federations, in countries as diverse as Australia, Ethiopia, Germany, India and Switzerland (Lago et al. 2022; Tremblay 2023). Public funds are transferred to poorer jurisdictions as unconditional grants, the amounts of which are calculated as a function of the difference between those jurisdictions’ per-capita tax bases and the federation average. Transfers can originate either from the federal government (vertical equalization) or from same-level sub-federal jurisdictions with above-average per-capita tax bases (horizontal equalization).

The main attraction of such systems is that they are less prone to strategic manipulation than equalization based on sub-federal revenues or tax rates (Boadway 2006). When transfers are conditioned on revenues, for instance, sub-federal jurisdictions have a direct incentive to lower their tax rates, as any static revenue loss is compensated by the equalization scheme. With tax-base equalization, incentive effects also exist, but they are only indirect.
In tax-competition models with suboptimally low equilibrium tax rates, tax-base equalization has been shown to raise equilibrium tax rates, and thus public expenditure levels, back toward the optimum (Köthenbürger 2002; Bucovetsky and Smart 2006; Gross 2021). In that sense, tax-base equalization increases both interregional equity and overall efficiency. This result, however, is specific to modelling choices. In models without tax-base mobility but with distortionary taxation (Smart 1998), as well as in models with revenue-maximizing “Leviathan” sub-federal governments (Köthenbürger 2005), tax-base equalization can worsen the distortions and be welfare-dominated by the fully decentralized equilibrium.

While the welfare implications of fiscal equalization schemes will inevitably be highly context dependent and nearly impossible to pin down rigorously, a body of empirical research exists that shows consistently how tax-base equalization pushes up equilibrium sub-federal tax rates (Buettner 2006; Egger et al. 2010; Holm-Hadulla 2020; Buettner and Krause 2020). Researchers typically look for some quasi-exogenous changes to equalization rates. The main analysis then consists of comparing the evolution of tax rates in affected and non-affected jurisdictions before and after the change in equalization rates. These studies find that higher equalization rates increase taxes.

Another welfare-relevant but less researched aspect of fiscal equalization is its implications for the spatial distribution of economic activity. Boadway (2006) pointed out that agglomeration economies might not be fully realized when equalization incentivizes production factors to remain in lower-productivity jurisdictions. A first formal evaluation of this proposition was conducted by Albouy (2012), who concluded that fiscal equalization in Canada leads to locational inefficiencies that reduce national income by 0.4% annually. Recent research, building on spatial general equilibrium models, has shown heterogeneous sub-federal taxes to be a source of significant spatial
misallocation (Fajgelbaum et al. 2019). One might conjecture that tax-base equalization, by compressing the distribution of sub-federal tax rates, would mitigate this misallocation. It would be interesting to incorporate fiscal equalization explicitly into such a framework.

The implementation of fiscal equalization raises policy issues as well. An important practical issue is the definition of the tax base to be equalized. For the system to work equitably, definitions of the tax base should be uniform across the federation, and jurisdictions should have no leeway to manipulate their reported tax bases. It has been documented that fiscal equalization acted as a catalyst for the harmonization of the laws and regulations delineating the relevant tax bases in Swiss cantons (Brülhart et al. 2023). An important practical as well as theoretical issue is how broadly to define the tax base that is considered in the equalization formula, and how to weight different elements of that tax base. Empirical evidence shows that equalization affects not only the taxation of the considered tax bases but also that of tax bases that do not enter the equalization formula (Miyazaki 2020). More work on this issue would be useful, to inform choices on the optimal breadth and composition of the aggregated tax base used for calculating equalization transfers.

Another issue of considerable practical importance is exogenous revenue needs. Tax-base equalization schemes are typically complemented by separate transfer programs targeted at jurisdictions deemed to have particularly high revenue needs for reasons unrelated to their choice of tax rate. In Switzerland, 15% of transfers are targeted at cantons with above-average spending needs.

10.2 Vertical transfers: General-purpose grants

Grants paid by an upper-level government to lower-level governments are a standard feature of fiscal equalization schemes and are probably more common overall than horizontal
transfers among same-level government. Such vertical grants can be non-earmarked, general-purpose in nature, or they can be paid out as a top-up to certain items of lower-level spending (“matching grants”).

General-purpose vertical grants have been studied primarily through the lens of the “flypaper effect”. It had been observed across a wide range of contexts that fiscal transfers and windfalls tend to be used almost entirely for additional public spending, while comparable shocks to jurisdiction-level private income are rarely associated with higher public spending: the money sticks where it hits (Inman 2009). In a frictionless, perfect-information democratic system with homogeneous agent-voters, the source of the shock should not matter.

One explanation of the phenomenon is imperfect politics: self-serving sub-federal politicians find it easier to hold on to funds received from outside than to raise taxes on their electorate (Courant et al. 1979; Inman 2009). The flypaper effect may have behavioral roots in fiscal illusion, as voters do not consider public and private funds as fungible (Hines and Thaler 1995), conducting instead some form of mental accounting (Becker et al. 2020). It is, however, also possible to explain the flypaper effect in settings with benevolent governments and rational voters, such as when taxation is distortionary and thus costly to the local population (Dahlby 2011), or where changes in grants are capitalized into local housing prices (Allers and Vermeulen 2016).

From a normative point of view, it is important to distinguish the mechanisms behind flypaper effects – bureaucratic appropriation versus benevolent optimization – empirically. An interesting step in that direction was taken by Gadenne (2017), who showed, based on Brazilian data, that sub-federal governments spend increases in local tax revenues in ways that benefit local populations more so than increases in general-purpose grant money. These results are consistent
with the political-economy explanation of the flypaper effect and caution against generous non-earmarked vertical transfers.

Supporting this view, Berset and Schelker (2020) show that a large increase in equalization payments to Swiss municipalities after a windfall increase in canton-level tax revenue triggered large and long-lasting municipal spending increases benefitting mainly public employees. For the same municipalities, Berset et al. (2023) document how unanticipated large increases in municipalities’ own tax bases tend not to be followed by large increases in spending. Local policy makers mainly smooth transitory shocks to their own tax base. Put simply, these findings suggest that, even in a high-income mature democracy, a general-purpose grant to sub-federal jurisdictions “sticks where it hits”, whereas own-tax revenue appears to be employed in a less self-serving manner by local governments.

10.3 Vertical transfers: matching grants

Federal transfers to sub-federal jurisdictions often take a more targeted, interventionist form than general-purpose equalization payments. Probably the most prevalent policy instrument of this type is a matching grant, whereby the central government tops up sub-federal expenditure on specific items in a statutorily given proportion. Medicaid and the SNAP program (food stamps) in the U.S. are classic examples. In theory, matching grants can be an effective instrument to correct for externalities generated by sub-federal policy choices (Figuieres and Hindriks 2002). Empirical studies show that matching grants are effective at increasing expenditure on targeted budget items (Leung 2022), but there is also a risk of sub-federal jurisdictions “gaming” the system by declaring non-targeted expenditures under the targeted budget headings (Baicker and Staiger 2005).
Agrawal et al. (2023) propose the concept of a “marginal corrective transfer” (MCT), representing the optimal federal transfer per dollar spent by a sub-federal jurisdiction on a particular public service or investment. The matching grant rate is then given by \( \frac{\text{MCT}}{1-\text{MCT}} \), where the MCT can also be negative. In theory, such grants could be optimally designed to internalize positive and negative expenditure spillovers and externalities arising from tax-base mobility and strategic policy interactions among sub-federal governments, as noted in Sections 4 and 6 above. Plausible ordinal comparisons can be made. Agrawal et al. (2023) compare six different policy areas using U.S. data, and they compute the highest MCT for K-12 education spending and the most negative MCT for state-level subsidies used in “bidding-for-firms” or tax credits.

The Agrawal et al. (2023) framework sets out the positive case for matching grants, but it also shows their limitations. In some cases, the external effects are so large that the optimal matching grant rate is infinite, implying that the public service in question would need to be centralized to correct the externalities at the margin. In cases of finite optimal grant rates, their determination still requires detailed and reliable estimates of relevant elasticities. Even if this information were available to governments, they might of course not act in the benevolent fashion but be driven by different self-serving or political motives.

Empirical evidence on matching grants is scant. Baker et al. (1999) have shown that capping matching grants to some Canadian provinces led them to lower their affected expenditures relative to a comparison group of provinces without such a cap. This finding confirms that matching grants, by lowering the local marginal cost of funds, “crowd in” local spending on the concerned items. To our knowledge, no empirically-grounded study exists attempting to quantify
the welfare effects of matching grants themselves, and of matching grants as compared to general-purpose grants.

11 Interjurisdictional cooperation and coordination

Recent policy debates concerning international cooperation and coordination of policy have focused on supra-national arrangements, e.g. through the European Union or OECD initiatives (Zodrow 2003; Devereux 2023). The empirical economics literature on supranational integration is comparatively small and, given the inevitably limited sample sizes, mostly theoretical (e.g. Alesina et al. 2005, Bordignon and Brusco 2006). The empirical study of cooperation and coordination has instead been focused on subnational agreements among municipalities or states, where larger samples and greater institutional variation can be leveraged for estimating causal relationships.

As an alternative to grants, intermunicipal cooperation is often used to internalize interjurisdictional externalities. Such cooperation can range from voluntary—perhaps informal—groups of municipalities that agree to commonly provide public services or set taxes, to arrangements that are compulsory or highly structured by federal law. Examples of voluntary cooperation include town twinning across international borders, the signing of treaties, or interjurisdictional compacts. On the more formal end, France recently required all towns to be a member of an establishment for intermunicipal cooperation (EIMC, see Tricaud 2022). Intermunicipal cooperatives are also common in Germany, France, Italy, Spain, and the US, although their formal implementation differs across countries (Hulst and Montfort 2007).

In the case of France, even though towns are required to cooperate, they decide with whom to cooperate. EIMCs only provide some public services—particularly, those that benefit from economies of scale, with municipalities retaining responsibility over other expenditures. EIMCs
may construct local development plans, provide waste/water/sanitation services, environmental services, road maintenance, and build libraries or other public facilities. In order to fund these common services, EIMCs decide how to delegate the taxation of four different local taxes between the cooperative and their municipalities. Under one regime, EIMCs and municipalities co-occupy all four tax bases. At the opposite extreme, the business tax is entirely delegated to the EIMC, with the municipalities solely taxing the other three bases. A hybrid regime that combines elements of both is also possible.

EIMCs can be viewed as a newly created tier of government, but with local governments choosing their partners and EIMC borders possibly changing over time, a tier that raises issues of coalition formation (Konrad and Schjelderup 1999; Burbidge et al. 1997; Agrawal 2023). Although EIMCs are designed to exploit economies of scale and mitigate horizontal tax competition, their creation may also create vertical externalities between municipalities and EIMCs sharing the same base. Even if the tax regime completely partitions the tax bases between the two tiers, such externalities may still arise due to complementarity or substitutability of the tax bases (Breuillé and Duran-Vigneron 2023).

How large should EIMCs be? As in Alesina and Spolaore (2003), the optimal size of jurisdictions trades off the benefits of size and the costs of heterogeneity. Further, Breuillé and Zanaj (2013) consider the incentives to create larger EIMCs via the merging of these higher-level tiers. EIMC mergers have three effects. First, the decline in the number of jurisdictions makes horizontal tax competition less intense while also reducing the distortions due to co-occupancy. Second, larger EIMCs can provide more public services at the same tax rate due to scale economies. Third, a larger EIMC better internalizes horizontal externalities.
Exploiting data from 36,000 municipalities in France and the recent creation of many new EIMCs, Breuillé, Duran-Vigneron and Samson (2018) show that EIMCs led to an increase in the total burden of local taxes, amounting to a 35% increase in the average tax bill. These increases were the greatest under the tax regime where EIMCs explicitly co-occupy all four tax bases. Although total tax rates increased, municipal tax rates fell, especially for the smallest municipalities. These results suggest that the upward pressure on higher-tier taxes is the dominant factor. Thus, despite scale-related gains in the provision of public services, intermunicipal cooperation is not necessarily a way to reduce the overall tax bill.

Another common form of cooperation is special districts. These districts, which typically cover several cities, can be viewed as allowing specific public services (water, sewers, parks, etc.) to be provided to optimal-size population groups, exploiting scale economies in cases where they exist. A related issue is economies of scope in public-good provision, which may cause several different public goods to be provided to a common population despite having different optimal group sizes (Brueckner and Lee 1991). Gains from economies of scope may justify this group-size compromise (examples are police and fire protection) but granting these groups the authority to raise tax revenue can result in special districts co-occupying municipal tax bases.

As to whether cooperation should be voluntary or compulsory, Tricaud (2022) exploits a French reform requiring municipalities to join an EIMC. She shows that holdout urban municipalities experienced increases in construction consistent with a reduction in NIMBYism, whereas rural holdouts saw declines in public services because the EIMCs centralized services such as libraries toward the more-populous municipalities. While Tricaud finds these effects for holdouts, she does not find similar effects for municipalities that voluntarily cooperated prior to
the law, suggesting that cooperation can impose added costs on some jurisdictions, perhaps related to their loss of political bargaining power with nearby communities.

11.1. Amalgamation and mergers

Other policies can achieve the same goal as EIMCs without creating an additional tier of government. For example, to achieve economies of scale in public good provision, municipalities can amalgamate or merge (see e.g. Hirota and Yunoue 2017). One key difference between mergers and EIMCs is that a merger requires the jurisdiction to surrender all tax and spending powers to the common government, while EIMCs generally allow municipalities to retain some local powers. The surrender of powers raises political-economy concerns: there may be an inefficiently low number of mergers, which stall because of disagreements over subsequent policies or even because of something as simple as a disagreement over the new municipality’s name.

12. Developing countries: A new direction for federalism research

Most economic fiscal-federalism research has focused on issues arising in high-income countries such as the US, Germany, Canada and Switzerland. However, large fiscal federations are found among developing economies, including India and especially China, where “market preserving federalism” (Weingast 1995) has been linked to economic success (Lü and Landry 2014). Gadenne and Sinhal (2014) show that, while developed countries have been centralizing when viewed from a centuries-long time scale, many developing nations attained formal state capacity at the central level first, and decentralized later. Reasons for the historical primacy of the central government in many developing countries include their colonial past, revenue sources such as natural resources and foreign aid lending themselves to centralized administration, and high inequality that requires stronger redistribution (Gadenne and Singhal 2014).
International organizations such as the IMF have long emphasized the potential advantages of greater fiscal decentralization in developing nations. The World Bank’s “rationale for decentralization” consists of three elements: “greater voice and choice to citizens”, local governments responding “dynamically to communities”, and “matching local needs and preferences with patterns of local public expenditure” (World Bank 2013). These themes match those discussed in Section 2, even though it has been argued that the conditions for Tiebout’s model do not hold in many low-income countries.

Academic research on this topic, as a result, has taken a broader political-economy perspective, taking account of “Leviathan” public sector motives and principal-agent problems—termed in a “second-generation theory of fiscal federalism” (Qian and Weingast 1997; Oates 2005). The “Leviathan-taming” effect of interjurisdictional competition in decentralized nations has been widely studied in high-income settings (Hoyt 1991; Edwards and Keen 1996; Brülhart and Jametti 2019). Interpreting government rent-seeking in the developing world as corruption (Brueckner 2000c), the Leviathan-taming effect of decentralization then becomes a corruption-taming effect, as shown by Arikan 2004.

But other researchers have strived to tailor models of fiscal federalism specifically to developing-country economic and political realities. On the one hand, it has been argued that the mechanisms at the heart of classic models of fiscal federalism are not “practical concerns” in many developing countries (Bardhan and Mookherjee 2006: 102). Competitive pressures on local governments arising from interjurisdictional mobility are less relevant in countries where “voting with one’s feet” over differences in public policies is costlier, and where spending autonomy typically is decentralized more than taxing rights (Gadenne and Singhal 2014; Bardhan and Mookherjee 2005).
Theoretical analyses by Bardhan and Mookherjee (2000, 2005, 2006) have also emphasized two alternative distortions that are arguably more relevant to developing nations: corruption and elite capture. In those models, corruption is primarily associated with federal-level bureaucrats tasked with regional-level policy delivery and subject to imperfect monitoring by the central government. Conversely, disproportionate representation of elite interests through lobbying and favoritism is modeled as being more prevalent in sub-federal politics than at the federal level, because local elections may be less contested and local special interests more easily organized. Bardhan and Mookherjee (2005: 679) write that with fiscal decentralization, “economic corruption tends to be replaced by political corruption.”

In those settings, new tradeoffs appear. For instance, anti-poverty programs may be more efficient under decentralization, because less is wasted through corruption. Yet, decentralization may also make those programs less well targeted to the neediest regions, as elite capture may be stronger in poorer states (Bardhan and Mookherjee 2005). The empirical evidence on this topic, however, is split (Galasso and Ravallion 2005; Balán et al. 2022). Similarly, sub-federal autonomy over taxation and spending will be regressive to the extent that local politics are captured by local elites. In short, while fiscal federalism in high-income countries may favor rich individuals because they are more mobile, it could favor them in low-income countries because they are better able to tilt local politics in their favor.

The early empirical literature in this area consisted of descriptive case studies (Bardhan 2002). Fisman and Gatti (2002) took a more ambitious approach by regressing an index of perceived corruption on a measure of spending decentralization, instrumented with the country’s legal origin. They found a robustly negative association between decentralization and corruption, a result confirmed by Arikan (2004). Similarly, Fan et al. (2009) regressed a firm-level measure
of experienced corruption on various decentralization measures in data for a cross-section of 80 countries in 1999-2000. However, they instead found that greater vertical and horizontal fragmentation of fiscal federations was associated with a higher incidence of bribery. In a quasi-experimental field study of decentralized versus centralized irrigation-management systems in Pakistan, Jacoby et al. (2021) found that decentralization to local farmer organizations was associated with greater theft and more unequal distribution in favor of the rich (large landowners). This finding, consistent with the elite-capture view of decentralization in developing countries, is reinforced by Reinikka and Svennson’s (2004) demonstration that that school districts in Uganda only receive 13% of the grants allotted to them, with the bulk captured by local officials and politicians. In contrast, Basurto et al. (2020) found that rural subsidy programs in Malawi were more efficiently targeted when administered by local chiefs than when distributed by a centralized mechanism, an opposite finding consistent with an informational advantage from decentralized policy-making.

Our reading of this literature is that the “decentralization optimism” dominating much of development economics thinking in the 1990s and early 2000s has been dampened by subsequent research, resulting in actual recentralization in some countries (Malesky, Nguyen and Tran 2014). Although the Tiebout model discussed in Section 2 has proven to be amazingly resilient over the last sixty years, the lessons learned from developing countries highlight limitations of the classic analysis, highlighting future avenues of research that take more of a political-economy approach. The problem is that standard criteria for judging the optimal degree of decentralization in high-income countries—interregional mobility, interregional externalities, and regional preference

5 For a comprehensive review of empirical evidence on intracommunity capture and corruption in developing countries, see Mookherjee (2015).
heterogeneity—are less central to corresponding analyses in low-income countries. In developing economies, distributional and political-economy issues tend to loom larger, and those considerations have been shown to be more favorable to centralized tax and spending decisions. Obviously, however, these are vastly simplified generalizations, and the literature has also shown that potential benefits from decentralization depend very much on institutional and economic contexts as well as on the precise nature of the decentralization.6

13. Federalism and the pandemic

The Covid-19 pandemic that erupted in early 2020 brought some of the advantages and drawbacks of fiscal decentralization into sharp focus. Control of an infectious disease touches on many issues that are familiar to fiscal-federalism researchers. Spatially unequal disease prevalence implies spatially unequal optimal policy responses, reminiscent of Tiebout-type differentiation. Quantitative analyses calibrated on U.S. data have concluded that a regionally targeted approach is more efficient than a uniform nationwide policy (Crucini and O’Flaherty 2020; Hoover and Toxvaerd 2022). On the face of it, this result would suggest that delegation of mitigation policy to the sub-federal level is optimal, although debates about whether the response to the pandemic should be centralized or decentralized were apparent within many countries and across the EU.

However, such measures, when taken by a particular sub-federal jurisdiction, generate public-health and economic externalities for other jurisdictions. For example, mitigation through

6 For example, even though decentralization in developing countries often primarily concerns spending decisions (Gadenne and Singhal 2014), it has been shown that full decentralization that includes taxing rights can have fundamentally different implications (Bardhan and Mookherjee 2006).
stay-at-home orders or business closures will yield a positive public-health externality for neighboring jurisdictions by reducing local virus circulation. There could also be negative externalities if mitigation measures in one jurisdiction were to shift contact-intensive interactions toward neighboring jurisdictions. Elenev et al. (2021), using U.S. county-level data, conclude that the former beneficial effect dominated. This finding implies that, with purely decentralized decision making, mitigation policies would be underprovided, as they yield positive externalities but impose “private” economic costs on the jurisdiction in question (Rothert 2021). This conclusion could imply that the optimal assignment of mitigation policy is to the federal government.

Such a view is supported by calculations of Renne et al. (2020), showing that, during the first pandemic wave, over two-thirds of U.S. deaths could have been prevented if policies that mirrored those of the earliest and strictest states had been adopted at the federal level. But when the costs of constraining economic activity, which vary regionally, are considered together with spatially heterogeneous disease prevalence, centralized policy-making that is regionally targeted then appears most attractive (Crucini and O’Flaherty 2020; Hoover and Toxvaerd 2022). Furthermore, in the context of mask mandates, centralized policy making has also been found to be more effective than decentralization (Seegert et al. 2020). Given that spatial externalities are at the heart of an epidemiological crisis, the case for centralized but regionally targeted decision-making seems strong.7

7 The related literature on disaster policy shows when the national government implements ex post disaster relief, the policy will create a moral hazard for subnational governments regarding disaster
Empirically, the pandemic experience offers valuable within-country policy heterogeneity, allowing researchers to gauge the effect of different mitigation measures. This literature is vast and still growing. Panel variation across U.S. states has, for instance, been used to estimate the effect of mask mandates on infections and deaths, the effect of lockdowns on economic activity, and the effect of policy stringency and fiscal policy on mental health. In this respect too, research on decentralized nations other than the U.S. could offer useful complementary insights.

14. Future of federalism

Recent decades have witnessed an increase in fiscal decentralization, as the sub-national share of total tax revenues has increased in many countries. With countries becoming more decentralized, intergovernmental grants have consumed a greater share of upper-level government spending. In the U.S., for example, federal grants to states and localities have steadily risen from 1% of GDP in the 1950s to 5.6% of GDP in 2021 (Clemens and Veuger 2023). In the United Kingdom, despite devolution to Scotland, Wales and Northern Ireland in the 1990s, experts continue to advocate further decentralization (e.g. Lockwood 2013). The rise of federalism is not just confined to OECD countries, with federalism suggested as a tool for economic growth, the reduction of corruption in developing countries (Bardhan and Mookherjee 2006), and a solution to federal gridlock/polarization. But, at the same time, recent economic and technological trends such as globalization, climate change, infectious diseases, and rising income inequality pose threats to federalism.

________________________
prevention, such that national mandates on ex ante prevention policies are necessary (Wildasin, 2011).
First, increased political polarization and the marginalization of some constituencies has led to increased fragmentation in many countries, as those constituencies seek to obtain local resources via the formation of new jurisdictions (Grossman and Lewis 2014). As U.S. states and localities have sought to strengthen their policy autonomy, political polarization and legislative gridlock within the federal government has left the central government unable to respond with new limits on sub-national jurisdictions (Stark 2021). Thus, polarization suggests a tradeoff in federal systems: a beneficial increase in policy diversity at the local level is offset by a lack of policy uniformity and simplicity, perhaps resulting in substantial compliance costs for multijurisdictional economic activity. Similarly, global issues such as climate change and infectious-disease control perhaps require centralized reforms, but in the absence of a national consensus, localities implement diverse policies.

Second, rising inequality in income and wealth poses challenges for fiscal federalism as the spatial concentration of top earners and wealth potentially creates inequities in fiscal capacity across jurisdictions. While many countries have extensive fiscal equalization systems, they may be unable to cope with growing regional disparities in income and wealth. Furthermore, with the rich and the poor increasingly separated across space, the scope for local income redistribution may narrow, posing new challenges for redistributive policy and the welfare state as implemented by sub-national governments.

Finally, globalization and technological change have reduced national governments’ ability to control the flow of people, goods, and businesses. Rodrik (1997) argues that the resulting growth of diversification within nations can lead to a backlash of protectionism and anti-immigrant sentiment, as seen today. As globalization has weakened the authority of national governments to set regulatory, trade and immigration policies, voters have demanded more accountability through
local policy, a process Boadway and Shah (2009) nickname “glocalization”. Thus, globalization and technological change again have generated a tradeoff: increasing mobility creates a need for supranational institutions to resolve its effects while at the same time dampening the authority of national governments, with voters then viewing localities as the more accountable policymakers.
Terms and Definitions

Tiebout model

a model where individuals sort across communities according to their demands for public goods

spillovers

exist when a community’s public goods yield benefits or costs to residents of neighboring jurisdictions

horizontal tax competition

tax-setting behavior that takes account of the interjurisdictional mobility of taxed resources

horizontal fiscal externality

a jurisdiction’s revenue gain due to an inflow of a mobile resource resulting from another jurisdiction’s tax increase

partial fiscal decentralization

a decentralization arrangement in which the central government retains control over aspects of subnational taxation and public expenditure

economies of scale

a force that reduces the unit cost of a good or service as its output rises

fiscal federalism

a structure in which public goods or services are financed and provided by multiple levels of government

co-occupancy

when different levels of government levy taxes on the same tax base

vertical fiscal externality

a jurisdiction’s revenue loss resulting from a tax increase by another level of government taxing the same base

vertical tax interactions

tax-setting behavior by one level of government that strategically responds to that of other levels
assignment problem

debate over which level of government should be assigned particular spending responsibilities or taxing powers

diagonal fiscal externality

a local jurisdiction’s revenue gain resulting from a tax increase by a neighboring state

intermunicipal cooperation

occurs when a subset of neighboring municipalities agree to jointly finance and provide public goods or services

vertical equalization

payments from the central government to equalize budgetary means across sub-national governments

horizontal equalization

payments among sub-national governments to equalize their budgetary means

tax-base equalization

horizontal equalization based on a formula that considers only jurisdictions’ tax bases

matching grants

co-payments from the central government to top up sub-national public expenditures at a rate proportional to sub-national spending

flypaper effect

tendency for governments to use intergovernmental grants for additional expenditures, not for tax cuts

second-generation fiscal federalism

theories of fiscal federalism that take into account the incentives of political actors

elite capture

when groups of individuals extract disproportionate advantages through the political process
References


54


