A Dynamic Model of Fiscal Decentralization and Public Debt Accumulation

Si Guo
Yun Pei
Zoe Xie
Abstract

This paper develops a dynamic infinite-horizon model with two layers of governments to study theoretically and quantitatively how fiscal decentralization affects local and central government debt accumulation and spending. In the model, the central government makes transfers to local governments to offset vertical and horizontal fiscal imbalances. But the anticipation of transfers lowers local governments’ expected cost of borrowing and leads to overborrowing ex ante. Absent commitment, the central government over-transfers to reduce local governments’ future need to borrow, and in the equilibrium both local and central debts are inefficiently high. Consistent with empirical evidence, when fiscal decentralization widens vertical fiscal imbalances, local governments become more reliant on transfers, and both local and central debts rise. Applied to Spain, the model explains 39 percent of the rise in total government debt when the vertical fiscal imbalances widened during 1988–1996, and 18 percent of the fall in debt when the imbalances narrowed during 1996–2006.

This paper is a product of the Office of the Chief Economist, South Asia Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at sguo@imf.org, yunpei@buffalo.edu, and lxie@worldbank.org.
A Dynamic Model of Fiscal Decentralization and Public Debt Accumulation*

Si Guo  
International Monetary Fund

Yun Pei  
University at Buffalo

Zoe Xie†  
World Bank; Federal Reserve Bank of Atlanta

Keywords: Fiscal decentralization, Public debt, Time-consistent policy, Soft budget constraint

JEL classifications: E61, E62, H74

*This paper was previously titled “Decentralization and Overborrowing in a Fiscal Federation.” We would like to thank Lue Eyraud and Lusine Lusinyan for sharing their data, and thank Gustavo Leyva for his help understanding some literature in Spanish. We thank audience and discussants at Atlanta Fed, Tsinghua Macro Workshop, SED, Econometric Society (Davis), EEA, Fed Macro & International Meetings, Philadelphia Macro Workshop, and the IADB for helpful comments. This paper has benefited from discussions with Manuel Amador, Jinhui Bai, Huixin Bi, Toni Braun, Lei Fang, Dirk Foremny, Grey Gordon, Jonathan Heathcote, Roozbeh Hosseini, Anastasios Karantounias, Karen Kopecky, Federico Mandelman, Fernando Martin, Victor Rios-Rull, Juan Rubio-Ramirez, Philippe Wingender, and Kei-Mu Yi. The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, IMF management, the Federal Reserve System, the Federal Reserve Banks, the IBRD/World Bank or its executive directors.

†Emails: Guo (sguo@imf.org). Pei (yunpei@buffalo.edu). Xie (xiexx196@gmail.com; lxie@worldbank.org).
1 Introduction

In many countries, revenue collection power is largely held by the central government, while spending responsibilities are disproportionately assigned to local governments.\(^1\) Because of this asymmetry—known as “vertical fiscal imbalances,” local governments’ own revenue often falls short of their spending, and the gap is filled by intergovernmental transfers from the central government and local governments’ borrowing. Existing empirical findings suggest that widening vertical fiscal imbalances in a country are associated with worsening aggregate fiscal performance, such as faster public debt accumulation (see e.g., Rodden 2002, Eyraud and Lusinyan 2013). As an illustration, Panel A of Figure 1 shows the gaps in local government finances for three OECD countries, and Panel B shows that total (central and local) government debt tends to rise faster when vertical fiscal imbalances widen.\(^2\)

The theoretical literature on the soft budget constraint problem offers one channel for this relationship: rising vertical fiscal imbalances worsen local governments’ soft budget constraint problem and lead to rising local government debt (Boadway and Shah 2007, Chapter 5).\(^3\) But because this literature uses mostly static or two-period models, it does not have anything to say about central government debt,\(^4\) which accounts for the majority of government debt changes in the data.\(^5\) Even less is known about the size of the effect of vertical fiscal imbalances on government debt accumulation.

In this paper, we quantify the effect of vertical fiscal imbalances on total government debt accumulation by developing an infinite-horizon model. The infinite-horizon is a valid choice because it allows us to consider central government debt alongside local government debt. The model delivers a novel result in the equilibrium: the central government over-transfers to local governments, to the extent that the marginal utility from local government spending is smaller than the marginal utility from central government spending. Applying the model to fiscal decentralization, wider vertical fiscal imbalances make local governments more reliant on transfers and exacerbate the soft budget constraint problem.

---

\(^1\)We use “central government” to refer to the national government in a country, and use “local government” to collectively refer to the subnational government such as a regional, provincial, or municipal government.

\(^2\)Appendix A includes plots for 25 other OECD countries and shows that changes in total government debt comove with vertical fiscal imbalances in a large set of countries over time.

\(^3\)In their survey of the soft budget constraint concept, Kornai et al. (2003) define soft budget constraint as the phenomenon that arises when “one or more supporting organizations (S-organizations) are ready to cover all or part of the deficit” of an organization that has a budget constraint (a BC-organization). In our setting, local governments are the BC-organizations, and the central government is the S-organization.

\(^4\)In a two-period model, central government debt is zero at the end of the second period by default. This means that even though it is possible to model central government borrowing between the first and second periods, a two-period model does not allow the accumulation of debt over time.

\(^5\)For example, for G-7 countries, the change in central government debt-to-GDP ratio during 1995–2015 ranges from 35% (Canada) to 133% (Japan); the change in subnational government debt-to-GDP ratio is much smaller, ranging from 2% (United Kingdom) to 13% (Japan).
Figure 1: Fiscal decentralization indicators and changes in total government debt
Panel A: Local governments’ own revenue (blue dash) and spending (red dotted).
Unit: percent of national GDP

Panel B: Vertical fiscal imbalances (blue dash, left) vs annual changes in total government debt (red dotted, right)

Note: Local governments’ own revenue excludes transfers from the central government. Vertical fiscal imbalances = (Local governments’ spending – Local governments’ own revenue)/Local governments’ spending × 100. Total government debt is the sum of central and local government debts. Annual changes in total government debt are calculated as the first difference of total government debt as percent of national GDP.

constraint problem, so local government debt rises. As transfers rise, the central government needs to borrow more, so central government debt also rises. We take advantage of fiscal decentralization reforms in Spain to quantitatively evaluate the impact of vertical fiscal imbalances on total government debt.

Our model features a fiscal federation with soft budget constraints on local governments and an inherent time-inconsistency problem. Local governments have proportionally more spending responsibilities than revenue. This asymmetry creates the vertical fiscal imbalances. Both local and central governments can borrow and spend. But some local governments have more borrowing autonomy and hence heavier debt service burdens than others, which creates horizontal fiscal imbalances among the local governments. The central government can make transfers to local governments to reduce the vertical and horizontal imbalances, through uniform lump-sum or debt-contingent transfers. The uniform lump-sum transfer is non-distortionary but it cannot address the cross-region horizontal imbalances. The debt-contingent transfer favors the local governments that are more heavily indebted and hence can potentially reduce horizontal imbalances ex post. But the anticipation of the debt-contingent transfer softens the local governments’ budget constraints and creates ex ante incentives for the local governments to overborrow. As such, there is an inherent time-inconsistency problem: ex ante the central government wants to promise lower future debt-contingent transfer to reduce overborrowing; ex post it wants to transfer more to reduce the vertical and horizontal fiscal imbalances.
Given the inherent time-inconsistency problem in the model, we focus on the time-consistent equilibrium. We characterize the Markov-perfect equilibrium, where the central government has no commitment over its future transfer, spending, and debt policies. We show two theoretical results in this equilibrium. First, the central government prefers debt-contingent transfer to uniform lump-sum transfer, because without commitment to future policies it does not consider the ex ante distortionary effect of debt-contingent transfer. As a result, the local governments overborrow relative to the efficient allocation benchmark—a common result in the soft budget constraint literature.

The second and more novel result is that in the time-consistent equilibrium, the central government over-transfers such that at the steady state, the marginal utility from local government spending is smaller than the marginal utility from central government spending. This is in sharp contrast with the common result in the literature, where the size of federal transfers is chosen to equalize the marginal utility of spending across different regions or different levels of governments. The reason lies in the infinite-horizon feature of our model. In a two-period model with soft budget constraints, the central government moves last and does not need to take into account the impact of its decisions on the future. It simply chooses the amount of transfers that equates the marginal utilities from different types of public spending. In an infinite-horizon model, the central government considers the impact of its decisions on local governments’ behaviors in the future. In particular, in our model, the central government today understands that local governments overborrow because of their anticipation of future debt-contingent transfers. Without the ability to commit to lower future debt-contingent transfers, the central government finds it optimal to over-transfer today to reduce local governments’ need to borrow. In the equilibrium, the over-transfer exacerbates ex ante overborrowing by local governments, and worsens the vertical fiscal imbalances.

Given the equilibrium properties, the model predicts that widening vertical fiscal imbalances are associated with faster government debt accumulation, which is consistent with past empirical findings. In the model, when vertical imbalances widen, local governments become more reliant on intergovernmental transfers to finance local spending. The anticipation of more transfers, and in particular more debt-contingent transfers, exacerbates local governments’ overborrowing. At the same time, central government debt rises to finance the higher transfers. In many countries, fiscal decentralization reforms tend to widen the vertical imbalances, because decentralization on the spending side often outpaces the revenue side. A number of international organizations have advocated for “balanced” fiscal decentralization (e.g., Lam et al. 2017; Sow and Razafimahefa 2017; Cibils and Ter-Minassian 2015). We formalize the idea and calculate the revenue decentralization needed to keep the observed expenditure decentralization “debt-neutral” or balanced.

---

6For example, in Sanguinetti and Tommasi (2004), the size of central-to-local transfers is chosen to equalize each region’s marginal utility from private consumption (which is comparable to local government spending in our model) and the marginal utility from central government consumption. In models without central government consumption, such as Goodspeed (2002a) and Kothenburger (2007), central-to-local transfer is chosen to equalize the marginal utility from public spending in each region, after accounting for the possible externality of public goods or voting preference.

7We also consider the case of (time-inconsistent) Ramsey allocation. We find that because the central government has commitment to future policies, it optimally under-transfers relative to the efficient allocation, which stands in contrast with the over-transfer result in the time-consistent Markov equilibrium.
To quantify the effects of vertical fiscal imbalances on government debt accumulation, we calibrate the model to Spain. Like many other countries, the local governments in Spain rely heavily on central government transfers, and the regions with higher debt tend to receive more transfers. The country has gone through several episodes of fiscal decentralization which changed the vertical fiscal imbalances. Empirically, as vertical imbalances widened, total government debt rose, which is consistent with the empirical relationship documented in the cross-country empirical studies in the literature.

Using the calibrated model, we find sizeable effects of changes in vertical fiscal imbalances on government debt accumulation in Spain. To take advantage of different fiscal decentralization reforms in our analysis, we look at two distinct episodes prior to the European debt crisis. In the first episode (1988–1996), faster decentralization of spending than revenue led to widening vertical imbalance and rising total government debt. In the second episode (1996–2006), decentralization of revenue caught up, so the vertical imbalances shrank, and total government debt fell. Quantitatively, we find that decentralization explains 39% of the debt increase in the first episode and 18% of the fall in the second episode.

Our analysis is closely related to the public economics literature on the soft budget constraint and common pool problems. Our paper complements this literature by using an infinite-horizon model. With the exception of Velasco (2000), most of the past literature uses static or two-period models. The infinite-horizon setup allows us to introduce central government debt and to study the dynamic interaction between current and future central government policies. An important implication of infinite-horizon is that absent commitment to future policies, the central government over-transfers to offset distortions from future central government policies, to the extent that the marginal utility from local government spending is smaller than (rather than equal to) the marginal utility from central government spending. Introducing central government debt also allows us to quantitatively study the impact of fiscal decentralization reforms on public debt accumulation.

A few papers in this literature also study the effects of fiscal decentralization on fiscal outcomes. The most related studies are Bellofatto and Besfamille (2018) and Garcia-Milà et al. (2002). Bellofatto and Besfamille (2018) use a three-period model to compare two different revenue arrangements for refinancing local projects: centralized versus decentralized tax revenue. Their focus is the impact of different fiscal arrangements on the provision of local public goods rather than on aggregate fiscal performance, and hence government debt is not modeled. Garcia-Milà et al. (2002) use a two-period model to study the impact of soft budget constraint and revenue decentralization on local government borrowing, when expenditure is already decentralized. In their model, whether or not the local governments can raise tax revenue affect their debt choices between the first and second period,

For other countries, Pettersson-Lidbom (2010) finds that Swedish local governments that received the highest numbers of discretionary transfers also had the largest accumulation of debt. Nicolini et al. (2002) show that the expectation of bailouts increases Argentina’s provincial governments’ incentives to run budget deficits. In Appendix B.2, we discuss anecdotal examples from Spain, Hungary, Italy, Germany, and several other countries where transfers are redistributive and approximately dependent on debt or deficit.

See, for example, Kornai (1986); Qian and Roland (1998); Velasco (2000); Sanguinetti and Tommasi (2004); Besfamille and Lockwood (2008); Bellofatto and Besfamille (2018); and Goodspeed (2016) for a review.
while central government is assumed to maintain a balanced budget. Compared to these papers, we use an infinite-horizon model to study the effect of fiscal decentralization on both local and central government debt accumulation. The inclusion of central government debt is important given that it is empirically more relevant in explaining the movements in total government debt. Additionally, our model allows us to study the full spectrum of spending and revenue decentralization in a quantitative exercise, instead of the two polar cases of full decentralization and full centralization.

Conceptually, the model shares common elements of the literature on time-inconsistency and time-consistent policies. We apply the idea of lack of commitment and policy distortion to the context of fiscal federalism. Importantly, we model the borrowing decisions of both central and local governments, and illustrate how central and local government debts interact with each other. In the model, to reduce local governments’ overborrowing, the central government without commitment over-transfers. The ability to borrow gives the central government the capacity to make even higher transfers than a government with a balanced budget. The higher transfer further increases local governments’ overborrowing, which leads to even higher total transfer and central government debt. As such, the interaction between central and local debt amplifies the aggregate effects of vertical imbalances on total debt.

The rest of the paper proceeds as follows: Section 2 presents the dynamic infinite-horizon model and characterizes the equilibrium and optimal policies. Section 3 calibrates the model to Spain. Section 4 contains quantitative applications to fiscal decentralization reforms and comparisons with other fiscal systems. Extensions and robustness checks to the baseline model are contained in this section. Section 5 concludes.

2 Model

In this section, we present an infinite-horizon model with two layers of governments, vertical fiscal imbalances and inter-governmental transfers. As a benchmark, we first characterize the allocations under a single consolidated government. We then consider the decentralized environment with two layers of governments making their own decisions separately. The allocations will depend on whether the central government has the ability to commit to its future transfers.

2.1 Model environment

We consider a small open economy. Time is discrete and infinite. There is a central government and a continuum of local regions indexed by $i \in [0, 1]$. Each region has one local government and a mass one of identical households. In each period, the representative household of a region receives endowment income $y$, pays tax $e$ and consumes the residual. The tax revenue is shared: $f$ for the local governments and $e - f$ for the central government. The sharing of spending responsibilities between the central and local governments is captured by a preference parameter $\theta$ to be elaborated.

---

10 See, for example, Klein, Krusell, and Ríos-Rull (2008); Martin (2011); Song, Storesletten, and Zilibotti (2012); Bianchi (2016); Karantounias (2017); Bianchi and Mendoza (2018).
later. Given these, each (local or central) government makes its spending and borrowing decisions each period, and the central government also determines transfers to local governments.

**Preferences.** The representative household living in region $i$ has the preference
\[
\sum_{t=0}^{\infty} \beta^t \left[ (1 - \theta)u(g_{i,t}) + \theta v(c_t) + w(y - e) \right]
\]
where $c_t$ is the per capita central government spending and $g_{i,t}$ is the per capita public spending of local government $i$, and

**Assumption 1.** The utility functions $u(\cdot)$ and $v(\cdot)$ are smooth, increasing and concave.

Both levels of governments are benevolent and utilitarian. Local government $i$ chooses $g_{i,t}$ and local debt issuance $b_{i,t+1}$ to maximize (1), subject to its local budget constraint. The central government maximizes the average utilities of households living in all regions, subject to a central budget constraint. We take household income $y$ and total government tax revenue $e$ as exogenous, so households’ utility from private consumption, $w(y - e)$, drops out from the central and local governments’ welfare maximization problems.

While the additive preference form in (1) over public goods provided by different levels of governments and the utility weight $\theta$ have been widely adopted in the literature,\(^{11}\) they can be micro-founded as follows when $u(\cdot)$ and $v(\cdot)$ are logarithmic and the elasticity of substitution is one. Suppose there is a unit mass of varieties of public goods, indexed by $\omega \in [0,1]$. Goods $\omega \in [0,\bar{\omega})$ are provided by the central government (e.g., national defense), and goods $\omega \in [\bar{\omega},1]$ are provided by the region’s local government (e.g., local roads and fire stations). Households in each region derive utility from the basket of public goods:
\[
U \left( \frac{1}{\sigma} \int_0^1 q(\omega) d\omega \right)
\]
where $q(\omega)$ is the quantity of good $\omega$. A decrease in $\bar{\omega}$ means more goods (or spending “lines”) are provided locally. This can be understood for example as the decentralization of education and health care spending in many countries. Under unitary elasticity of substitution between goods ($\sigma = 1$) and log utility, the preference becomes separable as in (1), with the parameter $\theta = \frac{\bar{\omega}^{1/\sigma}}{\bar{\omega}^{1/\sigma} + (1-\bar{\omega})^{1/\sigma}} = \bar{\omega}$ being the utility weight for goods provided by the central government. Appendix B.1 provides the details of the transformation.

**Vertical fiscal imbalances.** To make the model suitable for studying inter-governmental transfers, we assume that local governments’ share of total government revenue is not high enough to cover their spending obligations. This describes the vertical fiscal imbalances in many countries (Figure 1) and provides the motivation for central-to-local transfers. The precise mathematical assumption for vertical fiscal imbalances is given later by Assumption 4.

\(^{11}\)See, for example, Alesina and Tabellini (1990); Goodspeed (2002b); Besley and Coate (2003); Kothenburger (2007); Breuilé and Vigneault (2010).
Two types of local governments. We further introduce a cross-region (or “horizontal”) fiscal gap by modeling two types of local governments: the unconstrained (or type “n”) and the constrained (or type “o”). Unconstrained local governments $i \in [0, \mu]$, where $\mu \in (0, 1)$, have the full autonomy to borrow subject to an exogenous interest rate schedule. Constrained local governments $i \in (\mu, 1]$ face a constraint on borrowing, $b_{o,i,t} \leq B^o$. To make notations simpler, we assume $B^o = 0$ in this section, but results apply for any $B^o \geq 0$. When $\beta$ is small enough, which is guaranteed by Assumption 3 later, $b_{o,i,t} = 0$ always holds. Regions are otherwise identical. This difference in borrowing introduces a cross-region fiscal gap, as regions with more public debt also have higher debt service burdens.

We denote the average debt of unconstrained local governments by $b^n_t$, and $b^o_t$ for constrained local governments. Because all local governments of the same type are identical, we have $b_{i,t} = b^n_t$ for $i \in [0, \mu]$ and $b_{i,t} = b^o_t = 0$ for $i \in (\mu, 1]$. The nationwide (across all regions) local debt is then $b_t = \mu b^n_t$.

We introduce two types of local governments to create a link between local governments’ debt and their transfer income. If all regions are identical, the central government can simply use a lump-sum uniform transfer to replicate the allocation under a single consolidated government. With two types of regions, the central government has incentives to provide more transfers to the more indebted regions to offset the cross-region fiscal gap. This creates a positive correlation between the debt burden and transfer, which is the source of distortion in our model and is also supported empirically as we document in Section 3.1.

Government Debt. The central government and unconstrained local governments borrow from international financial markets, subject to exogenous and increasing interest rate schedules. The central government faces the gross interest rate schedule $S(d_t)$, where $d_t$ is the central government debt. The unconstrained local government faces the gross interest rate schedule $R(b^n_{i,t})$ for debt $b^n_{i,t}$. The separate interest rate schedules for the central and local governments make the analysis more transparent, and the distinction is not essential to the theoretical results. We make the following assumption on the interest rate schedules to ensure interior solutions.

**Assumption 2.** Both interest rate schedules $R(\cdot)$ and $S(\cdot)$ are increasing and convex functions, and there exist $\bar{B}, \bar{D} > 0$ such that $R(\bar{B}) = S(\bar{D}) = 1/\beta$.

Transfers. In each period, the central government optimally chooses the amount of transfers to each local government. At time $t$, for each region $i$, the only relevant individual state variable is its own

\[^{12}\]The lack of autonomy to borrow for some local governments can be understood either as the result of financial frictions or self-imposed fiscal discipline. In Appendix C and for all the proofs of propositions (Appendix E), we present a more general case where the constrained local governments face a constraint on borrowing $b_{o,i,t} \leq B^o$ with the exogenous debt limit $B^o \geq 0$.

\[^{13}\]In our setup, the difference in debt service burden across regions directly comes from the assumption that regions’ abilities to borrow differ. In Appendix F.1 we show an alternative setup, where both types of regions can borrow but with different interest rate schedules. We show that the different interest schedules can also lead to the cross-region difference in debt burden, and that similar theoretical results hold.

\[^{14}\]The theoretical results in the paper only require increasing interest rate schedules with ranges large enough to ensure an interior solution. The convexity assumption is one way to ensure the range of interest rate levels is large enough. In Appendix B.3, we provide two ways to motivate the increasing interest rate schedules.
debt stock $b_{i,t}$. The relevant aggregate state variables are the average debt of the unconstrained regions and the central government debt $(b^n_t, d_t)$.

In general, we can write the transfers received by region $i$ at time $t$, $T_{i,t}$, as a function of these state variables. For tractability we adopt a first-order approximation:

$$T_{i,t} = \alpha^0_t b_{i,t} + \alpha^1_t b^n_t + \alpha^2_t d_t + \alpha^3_t$$

where $\alpha^0_t, \alpha^1_t, \alpha^2_t$ and $\alpha^3_t$ are the (time-varying) coefficients chosen by the central government. Denote $\alpha^0_t = \tau_t$ and $T^n_t = \alpha^1_t b^n_t + \alpha^2_t d_t + \alpha^3_t$, this becomes

$$T_{i,t} = \tau_t b_{i,t} + T^n_t$$

which is a more general case for the affine tax/transfer functions used in the literature (e.g., Lucas and Stokey 1983, Aiyagari et al. 2002, Werning 2007, and Bianchi 2016).

Effectively, the central government has two transfer instruments: $T^n_t$ and $\tau_t$, which it decides upon knowing the aggregate states $(b^n_t, d_t)$. Each unconstrained region receives $\tau_t b^n_{i,t} + T^n_t$ and each constrained region receives $T^n_t$. Here $T^n_t$ can be interpreted as the uniform transfer identical across all regions, and $\tau_t$ is the rate at which the central government compensates for the debt service burden of unconstrained regions. Both instruments reduce the vertical fiscal imbalances between the central and local governments. Additionally, any positive $\tau_t$ also reduces the cross-region fiscal gap for given $b^n_t > 0$. As such, the transfer function can be interpreted as a form of redistributive transfer, whereby the central government has a tendency to transfer funds from fiscally rich to fiscally poor regions.\footnote{In Appendix B.2, we discuss anecdotal examples from Spain, Germany, and several other countries that transfers are redistributive and approximately dependent on debt or deficit. We also present empirical evidence showing that the Spanish central government implicitly takes into account local governments’ debt when it decides the amount of transfers.}

We further impose some lower bounds on both transfer instruments:

**Assumption 3.** $T^n_t$ and $\tau_t$ are subject to lower bounds:

(a) $T^n_t \geq \bar{T}$, where $\bar{T} \geq 0$ is a constant;

(b) $\tau_t \geq \bar{\tau}$, where $\bar{\tau} < 0$ is a constant and satisfies $\beta R(0)(1 - \bar{\tau}) < 1$.

The lower bound of uniform transfer, $\bar{T}$, captures that some basic locally-provided public goods are usually partially financed by federal transfers to ensure minimum quality and quantity (e.g., education and medical care). It is easy to see that $\bar{T}$ and local governments’ own revenue $f$ are substitutable: changing $(f, T)$ to $(f + \bar{T}, 0)$ would only decrease the equilibrium intergovernmental transfers by $\bar{T}$, without altering other elements of the equilibrium allocation. For our theoretical properties to hold, we only need $\bar{T}$ to be non-negative (that is, $T^n_t \geq \bar{T}$ can be relaxed to $T^n_t \geq 0$), so that the central government cannot tax away local governments’ own revenue. In the quantitative exercise, we calibrate $\bar{T}$ such that the simulated moment for total transfers matches the data.\footnote{In Appendix F.2, we relax Assumption 3(a). We instead assume that the central government faces an adjustment cost of changing $T^n_t$, and show that the theoretical properties of the model still hold.}
policy entails a zero \( \gamma_0 \) at the steady state. With the lower bound at \( \bar{\gamma} \) instead of zero, it is clear that the zero optimal \( \gamma \) is not a corner solution.

The following assumption states that the vertical fiscal imbalances are too large to be offset by the minimum uniform transfer alone when local government debt \( b^n = \bar{B} \) and central government debt \( d = \bar{D} \).

**Assumption 4. (Vertical fiscal imbalances)** \( f + \bar{T} \) is small relative to \( e \) such that

\[
(1 - \theta)u_g(f + \bar{T}) > \theta v_c(e - f - \bar{T} - \bar{D}(1 - \beta)),
\]

where \( \bar{D} \) satisfies \( S(\bar{D}) = 1/\beta \).

**Timing.** Figure 2 illustrates the timeline during period \( t \). Knowing the aggregate states \((b^n_t, d_t)\) determined in period \( t - 1 \) and central revenue \( e - f \), the central government moves first (Stackelberg leader) to choose its spending \( c_t \), transfer rate \( \gamma_t \), and uniform transfer \( \bar{T}_u \). The deficit is financed by new debt \( d_{t+1} \) issued at the interest rate \( S(d_{t+1}) \). An unconstrained local government \( i \in [0, \mu] \) with outstanding debt \( b^n_{i,t} \) receives its own revenue \( f \) and transfer \( \gamma_t b^n_{i,t} + \bar{T}_u \), chooses local public spending \( g^n_{i,t} \), and finances its deficit by issuing new debt \( b^n_{i,t+1} \) at the interest rate \( R(b^n_{i,t+1}) \). A constrained local government carries no debt and spends all its revenue on public spending \( g^n_{o,t} = f + \bar{T}_u \).

**Figure 2:** Timeline within period

<table>
<thead>
<tr>
<th>((b^n_t, d_t))</th>
<th>Central government</th>
<th>Local governments</th>
<th>((b^n_{t+1}, d_{t+1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>decides transfer policy ((\gamma_t, \bar{T}_u))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repays ( d_t ), borrows ( d_{t+1}/S(d_{t+1}) ), spends ( c_t )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repays ( b_{i,t} ), gets transfer ( \gamma_t b_{i,t} + \bar{T}_u )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>borrows ( b_{i,t+1}/R(b_{i,t+1}) ), spends ( g_{i,t} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( t + 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note on notation.** In our model, all regions of the same type are identical. As such, the many small local governments collapse into one representative constrained local government with population size \( \mu \) and one representative unconstrained local government with size \( 1 - \mu \). In the rest of the paper, wherever appropriate, we drop the subscript \( i \).

### 2.2 Efficiency benchmark: System with a consolidated government

Before we move to the equilibrium with two layers of government, we first characterize the allocation under a consolidated government that makes decisions for both central and local governments. The benevolent consolidated government maximizes the present value of the weighted average utility of all households. We use this “consolidated government allocation” as the efficiency benchmark.\(^{17}\)

**Definition 1.** Given \((b^n_0, d_0)\) and the interest rate schedules, the consolidated government’s problem

\(^{17}\)We assume that the consolidated government issues two types of debt: local debt and central debt. This assumption makes the consolidated government’s problem a closer benchmark to the equilibrium model with two layers of governments where both types of debt are issued (Sections 2.3-2.4).
consists of choosing a sequence of debt and spending \( \{b^n_{t+1}, d_{t+1}, g^n_i, g^n_o, c_t\}_{t=0}^{\infty} \) that solves

\[
\max_{\{b^n_{t+1}, d_{t+1}, g^n_i, g^n_o, c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \{ (1 - \theta)[\mu u(g^n_i) + (1 - \mu)u(g^n_o)] + \theta v(c_t) \}
\]

subject to the consolidated budget constraint

\[
c_t + \mu g^n_i + (1 - \mu)g^n_o + d_t + \mu b^n_t = e + \frac{d_{t+1}}{S(d_{t+1})} + \mu \frac{b^n_{t+1}}{R(b^n_{t+1})}
\]

and the constrained local governments’ no-borrowing condition, where \( g^n_i \) and \( g^n_o \) are the local spending of the unconstrained and constrained local governments, respectively.

Here we assume that governments do not internalize the slope of their interest rate schedules, which means that governments take their interest rates as given. As such, there are no derivatives of the interest rate functions \( R(\cdot) \) and \( S(\cdot) \) in the optimality conditions below.\(^{18}\) This modeling choice greatly simplifies notations, without changing the theoretical properties of the Markov equilibrium that we will discuss later. More importantly, when governments internalize interest rate schedules, there is a pecuniary externality, which potentially also generates overborrowing (e.g., Kim and Zhang 2012). We shut down this channel to highlight that overborrowing in our model is not a result of the pecuniary externality. In Section 4.3.1, we present an extension where governments internalize the slope of their interest rate schedules as a robustness exercise.

**Proposition 1.** The consolidated government’s optimal allocation satisfies

\[
(1 - \theta)u_g(g^n_i) = (1 - \theta)u_g(g^n_o) = \theta v(c_t)
\]

\[
\beta \frac{v(c_{t+1})}{v(c_t)} = \frac{1}{S(d_{t+1})}
\]

\[
\beta \frac{u_g(g^n_{t+1})}{u_g(g^n_t)} = \frac{1}{R(b^n_{t+1})}
\]

At the steady state, \( b^n_t = \hat{b}, d_t = \hat{d} \), and \( S(d_t) = R(b^n_t) \). Condition (3) indicates that the consolidated government achieves perfect resource sharing between the central and local levels and there is no cross-region difference in the quantity of locally-provided public goods. Under log utility, (3) implies that the central-local government spending ratio \( c_t / g^n_i = c_t / g^n_o = \theta / (1 - \theta) \). It is easy to see that the consolidated government allocation is efficient.

### 2.3 Distortion without a consolidated government

Without a consolidated government, each local or central government makes its own decisions. Central-to-local transfers distort the unconstrained local governments’ borrowing decisions, as their higher borrowings today lead to higher transfers tomorrow.

Given the aggregate states \( (b^n_i, d_i) \) and individual state \( b^n_{i,t} \), an unconstrained local government \( i \) chooses its spending \( g^n_i \) and debt \( b^n_{i,t+1} \) to maximize the welfare of its residents. Its problem written

\(^{18}\)More specifically, when a local government does not internalize its interest rate schedule, its marginal interest cost of higher \( b_i \) is simply \( R_i = R(b_i) \). In comparison, for a local government that internalizes the interest rate schedule, the marginal cost of higher \( b_i \) is \( R_i + \frac{2 R_i}{\theta b_i} b_i \), where the extra term represents the increased interest cost of \( b_i \).
recursive is
\[ W(b_{n,t}^n, b_{n+1}^n, d_{t}) = \max_{b_{n+1}^n, g_{t+1}^n} (1 - \theta)u(g_{t+1}^n) + \theta v(c_t) + \beta W(b_{n,t+1}^n, b_{n+1}^n, d_{t+1}) \] (6)

subject to the budget constraint
\[ g_{n,t}^n + b_{n,t}^n \leq f + \frac{b_{n+1}^n}{R(b_{n+1}^n)} + \tau_t b_{n,t}^n + T_u^u \] (7)

Taking interest rates and the central government's transfer policies \((\tau_t, T_u^u)\) as given, its optimal choice of \(b_{n,t+1}^n\) is characterized by the Euler equation,
\[ u_g(g_{n,t}^n) = \beta R(b_{n,t+1}^n)(1 - \tau_{t+1})u_g(g_{n,t+1}^n) \] (8)

Comparing (8) and (5), as long as \(\tau_{t+1} > 0\), the unconstrained governments \textbf{overborrow} relative to the local debt level chosen by a consolidated government. Given this distortionary effect of \(\tau\), the central government faces the following trade-off: On the one hand, the transfer through \(\tau\) provides more ex post debt relief to the unconstrained regions, who have higher debt service burdens than the constrained regions. On the other hand, all things equal, a positive \(\tau\) lowers local governments' borrowing cost and induces overborrowing. Because of this trade-off, there is a classic time-inconsistency problem. At time \(t\), the central government always wants to promise a low future transfer rate \(\tau_{t+1}\) (as well as \(\tau_{t+2}, \tau_{t+3}, \ldots\)) to reduce overborrowing ceteris paribus. When time \(t + 1\) comes, however, the central government would like to renege the promise of a low \(\tau_{t+1}\) and make more transfers to the unconstrained regions.

With this time-inconsistency problem, it matters whether the central government can commit to future policies. We first look at the equilibrium in which the central government does not have the ability to pre-commit to future policies \(\{\tau_t, T_u^u, c_t, d_{t+1}\}_{t > t_0}\) at any time \(t_0\). We then compare it with the equilibrium in which the central government can make commitment about future policies (the Ramsey solution). We also discuss ways to implement the consolidated government allocation regardless of commitment.

2.4 Equilibrium without commitment
2.4.1 Equilibrium definition
We define the equilibrium recursively (without time subscripts), using prime to denote next period values and dropping the subscript \(i\) for the local government. Given the beginning-of-period states of the economy \((b^n, d)\), the central government chooses spending \(c\), debt \(d'\), transfer rate \(\tau\) and uniform transfer \(T^u\) to maximize the welfare of its residents, subject to the unconstrained local governments’ Euler equation (8). We follow Klein et al. (2008) to define the Markov-perfect equilibrium where policy functions depend differentiably on current states \((b^n, d)\) and central government’s optimal policy is time-consistent.

For convenience, define the following functions of (representative) local and central government
spending, derived from their budget constraints

\[
\text{Unconstrained local: } G^u(b^n, b^n', \tau, T^u) = f + \frac{b^n'}{R(b^n')} - (1 - \tau)b^n + T^u \quad (9)
\]

\[
\text{Constrained local: } G^o(T^u) = f + T^u \quad (10)
\]

\[
\text{Central: } C(b^n, d, d', \tau, T^u) = e - f + \frac{d'}{S(d')} - d - \tau \mu b^n - T^u \quad (11)
\]

where as defined before \( \mu \) is the measure of unconstrained regions, and \( \mu b^n \) is the total amount of local government debt given constrained regions have zero debt.

**Definition 2.** A Markov-perfect equilibrium consists of a value function \( V \), central government’s policy rules \( \{\Phi^\tau, \Phi^T, \Phi^d\} \), and a policy function \( \Phi^b \) for the unconstrained local government’s debt, such that for all aggregate states \((b^n, d), \tau = \Phi^\tau(b^n, d), T^u = \Phi^T(b^n, d), d' = \Phi^d(b^n, d) \) and \( b^n' = \Phi^b(b^n, d) \) solve

\[
\max_{\tau, T^u, b^n', d'} (1 - \theta) \left[ \mu u \left( G^u(b^n, b^n', \tau, T^u) \right) + (1 - \mu) u \left( G^o(T^u) \right) \right] + \theta v \left( C(b^n, d, d', \tau, T^u) \right) + \beta V(b^n', d')
\]

subject to the representative unconstrained region’s Euler equation and a policy constraint,

\[
u_g \left( G^u(b^n, b^n', \tau, T^u) \right) = \beta R(b^n')(1 - \Phi^\tau(b^n', d')) u_g \left( G^u(b^n', \Phi^\tau(b^n', d'), \Phi^\tau(b^n', d'), \Phi^T(b^n', d')) \right) \quad (12)
\]

\[
T^u \geq \bar{T} \quad (13)
\]

and the central government’s value function satisfies the functional equation

\[
V(b^n, d) = (1 - \theta) \left[ \mu u \left( G^u(b^n, \Phi^b(b^n, d), \Phi^\tau(b^n, d), \Phi^T(b^n, d)) \right) + (1 - \mu) u \left( G^o(\Phi^T(b^n, d)) \right) \right] + \theta v \left( C(b^n, d, \Phi^d(b^n, d), \Phi^\tau(b^n, d), \Phi^T(b^n, d)) \right) + \beta V(\Phi^b(b^n, d), \Phi^d(b^n, d))
\]

In this equilibrium, the expected transfers tomorrow follow \( \tau' = \Phi^\tau(b^n', d') \) and \( T^{u'} = \Phi^T(b^n', d') \). Because each local government is infinitesimally small, it takes \{\( \tau, \tau', T^{u}, T^{u'} \)\} as given. However, when the central government decides \( \tau \) and \( T^u \) today, it understands that its choices will affect not only \( b^n', d' \), but also \( \tau' \) and \( T^{u'} \) through \( \Phi^\tau(b^n', d') \) and \( \Phi^T(b^n', d') \).

In the equilibrium, policy rules \( \Phi = \{\Phi^\tau, \Phi^T, \Phi^d, \Phi^b\} \) solve the central government’s problem each period, given the expectation that the central government’s choices in the next period and after will also follow \( \Phi \). In other words, if the central government today “announces” that \( \Phi \) will be implemented tomorrow and thereafter (even though it does not have commitment or credibility), then when tomorrow comes, it will indeed find \( \Phi \) optimal. This makes this equilibrium time-consistent.

### 2.4.2 Equilibrium characterization

The equilibrium can be characterized by conditions (12)–(13) and\(^{19}\)

---

\(^{19}\)Simplified notations, e.g., \( u^n_{gg} \) for \( u_{gg}(G^n) \), are used to make the optimality conditions more compact. We use prime to denote future variables, and subscript for partial derivatives, e.g., \( \Phi^T_{d'} \) denotes \( \partial \Phi^T(b^n', d')/\partial d' \). Appendix D.1 gives the derivation of these optimality conditions.

---

12
\[
\begin{align*}
\{\tau\} & \quad (1 - \theta)\mu u^n_g - \theta v_c = \lambda u^n_g \\
\{T^n\} & \quad (1 - \theta)[\mu u^n_g + (1 - \mu)u^n_g] - \theta v_c = \lambda u^n_g - \zeta \\
\{b^n\} & \quad (1 - \theta)\mu \frac{u^n_g}{R(b^n)} - \beta \left[(1 - \theta)\mu u^n_g (1 - \tau') + \theta v'_c \tau'\right] \\
& \quad = \lambda \left[\frac{u^n_g}{R(b^n)} + \beta R(b^n) (1 - \tau') u^n_g (1 - \tau')\right] + \lambda \Omega_{b^n} - \beta \lambda u^n_g (1 - \tau') + \beta \zeta' \Phi^T_{b^n} \\
\{d^n\} & \quad \theta \frac{v_c}{S(d')} - \beta \theta v_c' = \lambda \Omega_{d'} + \beta \zeta' \Phi^T_{d'}
\end{align*}
\]

where

\[
\Omega_x = \beta R(b^n) \Phi^T_x u^n_g - \beta R(b^n) (1 - \tau') u^n_g \left[b^n \Phi^T_x + \Phi^T_{x'} / R(b^n)\right], \quad x \in \{b^n, d^n\}
\]

and \(\lambda\) and \(\zeta\) are the Lagrange multipliers on constraints (12) and (13), respectively. The policy derivatives such as \(\Phi^T_{b^n}\) and \(\Phi^T_{d^n}\) capture the effects of today’s choices of \(b^n\) and \(d^n\) on future policies.

We can obtain some sharper characterizations of the Markov equilibrium at the steady state, given Assumptions 1-4.

**Lemma 1.** At the Markov equilibrium steady state, the unconstrained regions have lower public spending than the constrained regions: \(g^n < g^n\).

**Proof.** See Appendix E.1. \(\square\)

Because the unconstrained regions have larger debt service burdens than the constrained regions, the central government always wants to transfer more to the unconstrained regions than the constrained ones. This can only be achieved by transferring through \(\tau\). But because \(\tau\) encourages the unconstrained regions to borrow more, these regions end up with even larger debt service burdens. As a result, at the steady state, the spending gap between the two types of regions is not eliminated.

**Proposition 2.** At the Markov equilibrium steady state, the central government gives the minimum uniform transfer \(T^n = T\) and sets region-specific transfer \(\tau > 0\).\(^{20}\)

**Proof.** See Appendix E.3. \(\square\)

Intuitively, because the central government wants to transfer more to the unconstrained regions, it will always try to minimize the size of \(T^n\) and instead transfer through \(\tau\). The central government understands that \(\tau\) is distortionary as higher \(\tau_t\) leads to higher \(b^n_t\). If the central government can make policy commitment, it would like to pre-commit a low \(\tau_t\) in period \(t - 1\). But because the central government here does not have commitment, it takes \(b^n_t\) as given when it chooses \(\tau_t\) at time \(t\), and hence it does not take into account the ex ante overborrowing effect of \(\tau_t\) on \(b^n_t\). In other words,\(^{20}\)

\(^{20}\)It can be proved that Propositions 2 and 3 still hold when governments can internalize the slopes of interest rate schedules, i.e., when the optimality conditions (14)–(17) contain terms such as \(R_b^t\) and \(S_d^t\). We report the quantitative results under this alternate assumption in Section 4.3.1.
even though the central government knows that its choice of $\tau$ is distortionary ex ante, it does not internalize this distortionary effect, exactly because it cannot make an ex ante commitment about $\tau$.

The policy constraint $T^u \geq \bar{T} \geq 0$ is important for the equilibrium. Without it, the central government would want to do more cross-region redistribution by using a negative $T^u$ (i.e., a lump sum tax on all regions) to finance a higher $\tau$, which encourages local governments to borrow more ex ante, and leads to even higher $\tau$ and lower $T^u$. Equilibrium is thus either non-existing or exhibits an unrealistically high level of local debt.

**Proposition 3. (Over-transfer)** At the Markov equilibrium steady state, the average marginal utility of regional government spending is smaller than the marginal utility of central government spending, after adjusting for the utility weight $\theta$:

$$
(1 - \theta) \left[ \mu u^o_g + (1 - \mu) u^n_g \right] < \theta v_c
$$

(18)

**Proof.** Appendix E.2 provides the full proof. Here we offer a heuristic explanation for the proof. From condition (15), because $u^n_{gg} < 0$ and the Lagrange multiplier $\zeta \geq 0$, it is easy to see that the essence of proving Proposition 3 is to show that the Lagrange multiplier for the representative unconstrained region’s Euler equation (12) is positive: $\lambda > 0$. Condition (12) can be replaced with two inequalities:

\begin{align}
    u^n_g &\leq \beta R'(1 - \tau') u^n_g' \quad \text{and} \\
    u^n_g &\geq \beta R'(1 - \tau') u^n_g'
\end{align}

(19)  (20)

If we can show that only (19) is potentially binding around the steady state while (20) is redundant, we can come to the conclusion that $\lambda$ is positive around the steady state. To see why (20) is redundant, note that in the absence of (12), the central government would always want to choose allocations such that $u^n_g = \beta R' u^n_g'$, which is sufficient to ensure (20) holds for any transfer rate $\tau' > 0$. 

The central government’s over-transfer result in Proposition 3 is not a direct implication of local governments’ overborrowing. In the consolidated government benchmark, perfect resource sharing in condition (3) implies that the ($\theta$-adjusted) average marginal utility of local governments (“MULG”) exactly equals the ($\theta$-adjusted) marginal utility of the central government (“MUCG”). That is, the left-hand side of (18) would exactly equal the right-hand side. In the Markov equilibrium, because transferring through $\tau$ is distortionary, one may (mistakenly) think that the central government would under-transfer, to the extent that MULG > MUCG. To the contrary, Proposition 3 states that the central government over-transfers in the Markov equilibrium, to the extent that MULG < MUCG.

The reason behind the over-transfer result includes two layers. First, despite the distortionary effect of $\tau_t$, the central government does not have incentives to reduce $\tau_t$ from the level required for MULG = MUCG. This is because in the Markov equilibrium, the time-$t$ central government takes the local debt $b^n_t$ as given, and hence neglects the ex ante overborrowing effect of $\tau_t$ on $b^n_t$. Second and more importantly, the central government has an incentive to increase $\tau_t$ to be above the level
required for $\text{MULG} = \text{MUCG}$. This is because the marginal cost of local borrowing is higher for the entire fiscal federation $(R(b^n_{t+1}))$ than for the unconstrained regions $((1 - \tau_{t+1})R(b^n_{t+1}))$. This difference gives the central government an incentive to lower local debt $b^n_{t+1}$. But without commitment, the central government cannot promise lower future transfer rate $\tau_{t+1}$ to reduce the unconstrained regions’ incentive to borrow. Instead, it increases the current transfer rate $\tau_t$ to relax the unconstrained regions’ budget constraint at time $t$ and hence their need to borrow. The higher transfers are financed by lower central government spending (and higher central government debt), which explains why $\text{MULG} < \text{MUCG}$.

The over-transfer result of Proposition 3 is somewhat unique to our infinite-horizon environment. In a typical static or two-period model setting, the central government’s lack of commitment is modeled as the central government moving after local governments or private agents. In the last period of such models, as the last mover, the central government has no future (local or central) governments’ actions to take into account. That is, its optimization problem does not have local governments’ Euler equation (12) as a constraint. It thus sets the transfer in the last period to achieve the within-period perfect resource sharing, $\text{MULG} = \text{MUCG}$. With infinite horizons, however, the central government takes into account how its choice of $(\tau, T^n)$ affects future (central and local) governments’ choices. The impact of raising $\tau$ on local government’s future debt choice $b'$, which is captured by the right hand side of (14), drives the over-transfer result.

2.4.3 Effects of fiscal decentralization

How do fiscal decentralization reforms affect aggregate fiscal performance? In the model, a revenue decentralization (larger $f$) results in lower total government debt, whereas a spending decentralization (higher $1 - \theta$) leads to higher total government debt. We use numerical examples to help illustrate the effects of fiscal decentralization on the steady state debt levels. The examples are based on the calibration in Section 3.

A revenue decentralization increases local governments’ revenue $f$ and decreases the central government’s revenue $e - f$ one-to-one. This narrows the vertical fiscal imbalances. All else equal, the central government wants to reduce its transfers to local governments. Because the equilibrium...
uniform transfer $T^u$ is already at the lower bound (Proposition 2), the central government can only reduce transfers with a lower $\tau$. The lower $\tau$ in turn reduces the unconstrained regions’ overborrowing. As both $\tau$ and $b^n$ fall, total transfer $\tau b + T^u$ falls more than the drop in central government revenue. As a result, central government spending $c$ increases and debt $d$ falls. Total government debt $\mu b^n + d$ therefore also falls. Panel (a) of Figure 3 illustrates these changes numerically.

A spending decentralization increases local governments’ share of spending responsibilities $(1 - \theta)$. This widens the vertical fiscal imbalance. All else equal, the central government has greater incentives to make transfers. Because the central government is not willing to increase its uniform transfer $T^u$ above the lower bound (Proposition 2), it instead raises $\tau$. The higher $\tau$ increases the overborrowing incentive and raises the unconstrained regions’ debt $b^n$, as illustrated in Panel (b) of Figure 3. The higher $b^n$ widens the spending gap between the two types of regions and calls for an even larger $\tau$ to reduce the cross-region spending difference. The central government’s public spending $c$ falls as it has a smaller share of spending responsibilities (lower $\theta$). But the increase in transfer from larger $\tau$ and $b^n$ more than offsets the smaller spending, and the central government increases debt $d$ to meet the additional financing needs. As a result, total government debt $\mu b^n + d$ also rises.

**Figure 3:** Numerical illustration of fiscal decentralization in Markov equilibrium

(a) Effects of revenue decentralization

(b) Effects of spending decentralization

Note: Numerical illustration using the calibrated model. Details on calibration are in Section 3. Red star marks the calibrated 1996 economy. Black line plots how the model steady state changes as $f$ or $\theta$ changes.

The interaction between central and local debt amplifies fiscal decentralization’s impact on total debt. Wider vertical imbalances give the central government an incentive to increase the transfer rate $\tau$, which leads to a higher $b^n$. Because the central government can borrow, it can afford an even higher transfer rate through debt financing, which would not be possible if it had to maintain a balanced budget. The even higher transfer further exacerbates local governments’ overborrowing ex ante. A new steady state is reached when $b^n$ and $d$ increase to sufficiently high levels, such that the central
government becomes too financially stretched to further increase the transfers.

**Remark.** Before we end this section, we want to highlight that the only assumption we have made on the interest rate schedules $S(\cdot)$ and $R(\cdot)$ is that they are increasing and convex functions (Assumption 2). In particular, we do not require that the central government’s interest rate schedule $S(\cdot)$ is more favorable than local governments’ $R(\cdot)$. As such, the over-transfer result (Proposition 3) and the impact of fiscal decentralization are not because the central government wants to replace local debt with cheaper central debt to reduce net interest cost. Even if we do not allow the central government to borrow (i.e., $d_t \equiv 0$ for any $t$), the propositions in Section 2.4.2 still hold, and a revenue (spending) decentralization still reduces (increases) local government debt.\(^{24}\)

### 2.5 Equilibrium with commitment (Ramsey)

In the Ramsey problem, the central government chooses the entire sequence of policies and allocations at time 0, and sticks to the choices at any time $t \geq 0$. Because of this commitment to future policies, the Ramsey central government takes into account how time-$t$ policy affects local governments’ decisions at time $t - 1$ (as well as $t - 2, t - 3, \ldots$). This is the key difference from the Markov equilibrium. Formally,

**Definition 3.** Given initial debt positions $(b^u_0, d_0)$, the optimal (Ramsey) policy of a central government with commitment consists of a sequence of debt and transfer policies $\{\tau_t, T^u_t, b^u_{t+1}, d_{t+1}\}_{t=0}^{\infty}$ that solves

$$
\max_{\tau_t, T^u_t, b^u_{t+1}, d_{t+1}} \sum_{t=0}^{\infty} \beta^t \left\{ (1 - \theta) \left[ \mu u \left( G^n (b^u_t, b^u_{t+1}, \tau_t, T^u_t) \right) + (1 - \mu) u \left( G^o (T^u_t) \right) \right] + \theta v \left( C (b^c_t, d_t, d_{t+1}, \tau_t, T^u_t) \right) \right\}
$$

subject to the unconstrained regions’ Euler equation and transfer policy constraint for all time $t \geq 0$

$$
u_g \left( G^n (b^u_t, b^u_{t+1}, \tau_t, T^u_t) \right) = \beta R (b^h_{t+1}) (1 - \theta_{t+1}) u_g \left( G^m (b^h_{t+1}, b^h_{t+2}, \tau_{t+1}, T^u_{t+1}) \right)
$$

$$T^u_t \geq \bar{T}
$$

where the budget constraints of the central and local governments are implicit in the spending functions $C$, $G^n$, and $G^o$ as defined in (9)–(11).

Let $\beta^t \gamma_t$ and $\beta^t \zeta_t$ be the Lagrange multipliers on the two constraints above respectively. Written\(^{24}\)

The only change to the derivations in Section 2.4.2 is that Equation (17), the first-order condition with respect to $d'$, is replaced with $d' = 0$.\(^{24}\)
recursively, the first-order conditions of the Ramsey problem are\(^{25}\)

\[
\{\tau\} \quad (1 - \theta)\mu u^n_t - \theta \mu c = \gamma u^n_{tg} + \gamma^- R(b^n) \left[ u^n_t / b^n - (1 - \tau) u^n_{tg} \right] \tag{23}
\]

\[
\{T^n\} \quad (1 - \theta)\left[ \mu u^n_t + (1 - \mu) u^n_{tg} \right] - \theta v_c = \gamma u^n_{tg} - \gamma^- R(b^n) (1 - \tau) u^n_{tg} - \zeta^r \tag{24}
\]

\[
\{b^n\} \quad (1 - \theta)\mu u^n_{tg} - \beta [(1 - \theta)\mu u^n_{tg} (1 - \tau) + \theta v'_c \mu r']

= \gamma \left[ \frac{u^n_{tg}}{R(b^n)} + \beta R(b^n) (1 - \tau') u^n_{tg} (1 - \tau') \right] - \gamma^- R(b^n) (1 - \tau) \frac{u^n_{tg}}{R(b^n)} - \beta \gamma u^n_{tg} (1 - \tau') \tag{25}
\]

\[
\{d'\} \quad \frac{\theta v_c}{S(d')} - \beta \theta v'_c = 0 \tag{26}
\]

The distinctions between the Ramsey equilibrium and the equilibrium without commitment (Markov) can be seen, for example, by comparing (14) with (23). Both conditions characterize the direct effect of increasing \(\tau\). Because the Ramsey central government can pre-commit to future policies, it takes into account that an increase in \(\tau\) raises the unconstrained regions’ borrowing in the previous period. This “commitment effect” is captured by the \(\gamma^-\) term in (23), where \(\gamma^-\) is the shadow value of loosening the unconstrained local governments’ Euler equation in the previous period. By contrast, in the Markov equilibrium, the central government does not consider how its choice today affects local governments’ borrowings in the previous period, and so there is no \(\gamma^-\) term in (14).\(^{26}\)

This distinction gives rise to the potential time-inconsistency problem of the Ramsey allocation. At time 0, when the central government chooses the sequence of optimal transfer plan \(\{\tau_t, T^n_t\}_{t=0}^\infty\), it understands that the time-\(t\) policy \(\tau_t\) affects unconstrained local governments’ debt \(b^n_t\). However, this \(\tau_t\) may be suboptimal at time \(t\) after \(b^n_t\) is realized, because the effect of \(\tau_t\) on \(b^n_t\) is already foregone. In other words, the policy that is optimal before observing \(b^n_t\) may differ from the optimal plan after \(b^n_t\) is realized.

We show that the Ramsey equilibrium steady state exhibits under-transfer, in contrast to the over-transfer property in the Markov equilibrium steady state.

**Proposition 4. (Under-transfer)** At the Ramsey equilibrium steady state, the central government sets the distortionary transfer rate \(\tau = 0\) and the uniform transfer \(T^n > \bar{T}\). Additionally,

\[
(1 - \theta) \left[ \mu u^n_g + (1 - \mu) u^n_{tg} \right] > \theta v_c \tag{27}
\]

**Proof.** See Appendix E.4.\(^{27}\)
Here we give some intuition for Proposition 4. The Ramsey central government takes into account the overborrowing effect of $\tau$ (the term with $\gamma^-$ in condition 23). Proposition 4 shows that this additional cost is large enough to offset the incentive to use $\tau$ to close the cross-region fiscal gap. Note that $\tau = 0$ is not a corner solution as $\tau$ is bounded below by a negative value (Assumption 3) and not by 0. With $\tau = 0$, it is then obvious that the Ramsey central government cannot redistribute between the two types of regions. As long as there is a cross-region fiscal gap, local public spending in the two regions cannot be equalized ($g^n \neq g^o$), and the Ramsey solution is not efficient. It is worth noting that because the unconstrained regions overborrow when $\tau > 0$, using $\tau$ to transfer will only widen the cross-region spending gap, and the Ramsey central government internalizes that.

Note that (27) establishes that $\text{MULG} > \text{MUCG}$. That is, at the steady state, the Ramsey central government “under-transfers”: its optimal choice of uniform transfer $T_u$ is smaller than the level needed to achieve perfect resource sharing. To see why, note that a marginal increase in $T_u$ has three effects, captured by terms in the condition (24). First and foremost, it shifts resources from central to local governments (the left-hand side of 24). If this is the only effect from a higher $T_u$, then the optimal $T_u$ would achieve perfect resource sharing ($\text{MULG} = \text{MUCG}$). Second, a higher $T_u$ relaxes the minimum uniform transfer constraint $T_u \geq \bar{T}$, captured by the $\zeta^+$ term in (24). This effect is zero at the steady state ($\zeta^+ = 0$), because with $\tau = 0$ the vertical fiscal imbalances require a large $T_u$ which makes $T_u \geq \bar{T}$ non-binding. Third, for any marginal increase in $T_u$ at time $t$, the unconstrained regions want to smooth spending inter-temporally by increasing $b^n_t$ and decreasing $b^{n+1}_t$. These effects are captured on the right-hand side of (24) by the terms containing the shadow values $\gamma$ and $\gamma^-$ of relaxing the unconstrained regions’ Euler equation. At the steady state, the net effect of higher $T_u$ through the changes in local governments’ borrowings is $-\gamma u^n_{yy} + \gamma R(1-\tau)u^n_{yy}$ and is negative (i.e., a net social cost) when $\tau = 0$. Because of this additional social cost of higher $T_u$, the central government sets $T_u$ below the efficient level and hence $\text{MULG} > \text{MUCG}$.

**COROLLARY 1.** The Ramsey steady state local and central government debt levels are the same as in the consolidated government allocation, $b^n = \bar{B}$ and $d = \bar{D}$.

Because (26) is the same as (4), central debt at the Ramsey steady state is the same as under the consolidated government. It follows from $\tau = 0$ that (25) is identical to (5) at the steady state, and so the Ramsey steady state $b^n$ is also the same as the level under the consolidated government.

**Summary of theoretical results.** Table 1 summarizes the theoretical results.

### 2.6 Implementation of consolidated government allocation

As illustrated in Table 1, neither the Markov nor the Ramsey equilibrium can achieve the consolidated government allocation. In this section, we show that a prudential tax, combined with the
Table 1: Summary of Theoretical Results at Steady State

<table>
<thead>
<tr>
<th>Transfer policy</th>
<th>Consolidated government allocation</th>
<th>Equilibrium without consolidated government</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-commitment (Markov)</td>
<td>Commitment (Ramsey)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer policy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No overborrowing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debt</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-commitment (Markov)</td>
<td>Commitment (Ramsey)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal spending gap</td>
<td>None, $g^n = g^o$</td>
<td>None, $g^n = g^o$</td>
</tr>
<tr>
<td>Vertical spending gap</td>
<td>MULG = MUCG</td>
<td>MULG &lt; MUCG</td>
</tr>
</tbody>
</table>

Note: This table summarizes the theoretical results of Propositions 1–4, Lemma 1, and Corollary 1. MULG (MUCG) stands for the $\theta$-adjusted average marginal utility of local (central) governments: $\text{MULG} = (1 - \theta)[\mu u^n g + (1 - \mu)u^n o g]$; $\text{MUCG} = \theta v_c$. We assume governments do not internalize the slopes of interest rate schedules.

existing transfer instruments $(\tau_t, T^n_t)$, can achieve this efficient allocation. The optimal policy with this prudential tax is time-consistent, which means that the central government does not have incentives to deviate from it even without commitment to future policies. Intuitively, similar to Bianchi (2016), the prudential tax offsets the ex ante overborrowing distortion created by $\tau_t$, and allows the central government to use the transfer instruments $\tau_t$ and $T^n_t$ to exactly offset both the vertical and horizontal fiscal imbalances.

We endow the central government with an additional instrument: a prudential tax $\tau_t x_t$ on local government borrowing, to be implemented ex ante. The tax proceeds are rebated to unconstrained local governments in a lump sum fashion. Figure 4 illustrates the timeline with the additional instrument.

![Figure 4: Timeline with prudential tax](image)

Formally, the unconstrained local governments’ problem is similar as before, except that their budget constraint (7) becomes

$$g^n_{i,t} + b^n_{i,t} \leq f + \frac{(1 - \tau_t x_t)b^n_{i,t+1}}{R(b^n_{i,t+1})} + \tau_t b^n_{i,t} + T^n_t + T^n_{reb,t}$$

The lump-sum rebate $T^n_{reb,t} = \tau_t x_t b^n_{i,t+1} / R(b^n_{t+1})$ is equal to the total tax receipt (hence no $i$ on $b^n_{t+1}$) equally distributed among unconstrained regions, and is taken as given by local governments. Constrained local governments’ revenue and spending are the same as before, as they are not subject to
the prudential tax or receiving any rebate.

The central government’s Ramsey problem with prudential tax (R2) becomes

$$\max \{c_t, g_t^n, g_t^n, d_{t+1, \tau_t, T_t^u, \tau_t} = \sum_{t=0}^{\infty} \beta^t \left\{ (1 - \theta) \left[ \mu u(g_t^n) + (1 - \mu) u(g_t^0) \right] + \theta v(c_t) \right\}$$

subject to

$$\begin{align*}
(1 - \text{tax}_t) u_g(g_t^n) &= \beta R(b_{t+1}^n)(1 - \tau_{t+1}) u_g(g_{t+1}^n) \\
g_t^n + b_t^n &= f + \frac{b_{t+1}^n}{R(b_{t+1}^n)} + \tau_t b_t + T_t^u \\
g_t^0 &= f + T_t^u \\
c_t + \mu g_t^n + (1 - \mu) g_t^0 + d_t + \mu b_t^n &= e + \frac{d_{t+1}}{S(d_{t+1})} + \mu \frac{b_{t+1}^n}{R(b_{t+1}^n)}
\end{align*}$$

and $T_t^u \geq \bar{T}$. Equations (29)–(31) are the Euler equation and budget constraints of the representative local governments. Equation (32) is the resource constraint for the entire government sector. Notice that (30) comes from aggregating the budget constraints of all unconstrained regions (28), and is identical to the budget constraint without prudential tax because total tax receipt equals total rebate.

**Proposition 5.** Suppose the central government can implement a prudential tax $\text{tax}_t$ on local debt $b_{t+1}^n$ at time $t$, in addition to the transfer instruments $\tau_t$ and $T_t^u$. Then the Ramsey optimal policy that solves (R2) can achieve the consolidated government allocation and is time-consistent.

**Proof.** We show that the consolidated government allocation satisfies (29)–(32) with appropriately constructed sequence $\{t\hat{\hat{a}}_t, \hat{\tau}_t, \hat{T}_t^u\}$. Given $(b_0^n, d_0)$, the consolidated government allocation, denoted by $\{b_t^n, d_t, g_t^n, g_t^0, c_t\}$, automatically satisfies the resource constraint (32). Choose $\{\hat{\tau}_t, \hat{T}_t^u\}$ such that the local governments’ budget constraints (30)–(31) are satisfied

$$\begin{align*}
\hat{\tau}_t b_t^n + \hat{T}_t^u &= g_t^n + b_t^n - f - \frac{b_t^n}{R(b_{t+1}^n)} \\
\hat{T}_t^u &= g_t^0 - f
\end{align*}$$

Construct $\{t\hat{a}_t\}$ such that

$$t\hat{a}_t = \hat{\tau}_{t+1}$$

It is then straightforward to verify that Euler equation (29) becomes the same as the consolidated government’s optimality condition with respect to $b_{t+1}^n$ (condition 5). Therefore, the consolidated government allocation $\{b_t^n, d_t, g_t^n, g_t^0, c_t\}$ and the constructed policy sequence $\{t\hat{a}_t, \hat{\tau}_t, \hat{T}_t^u\}$ together solve this new Ramsey problem (R2). The time-consistency property follows immediately from the consolidated government allocation being efficient and hence time-consistent.

3 **Calibration to Spain**

In this section, we describe the calibration of our model to Spain, which we use for quantitative applications in the next section. We calibrate the model to data moments in Spain for two reasons.
First, Spain went through several well-documented fiscal decentralization reforms, and the decentralization on the revenue and spending sides was asymmetric. This allows us to test and quantify the model predicted relationship between vertical fiscal imbalances and total government debt accumulation (Section 2.4.3). Second, despite the fiscal decentralization reforms, Spain’s regional governments had little autonomy on revenue policies (e.g., deciding tax rates) but more autonomy to decide spending. This is consistent with our modeling choices of endogenous government spending and exogenous pre-transfer revenue. We focus on the period before the European debt crisis and Global Financial Crisis by excluding post-2006 data.  

3.1 Background: fiscal decentralization in Spain

Fiscal decentralization. Since the creation of 17 financially autonomous regions (the Autonomous Communities or ACs) in 1978, the funding of these regions has been periodically reformed. During the 1980s and 1990s, the country went through several fiscal decentralization reforms to gradually increase regional governments’ shares of tax revenue and spending responsibilities. On the spending side, changes in the national versus regional spending responsibilities were mainly driven by shifts in the provision of health care and education from central to the regions at different points in time. On the revenue side, some tax revenue items were ceded to regional governments to increase regional governments’ own revenue. From 1986 onward, the fiscal arrangements between the central and regional governments were renegotiated about every 5 years.

We treat the regional and sub-regional governments in the data as the local governments in our model. Panels a and b of Figure 5 show that the local governments’ pre-transfer revenue and spending as shares of total government revenue and spending both went up after the mid-1980s. Two observations are worth noting. First, the local governments’ share of total revenue was always smaller than their share of total spending, indicating vertical fiscal imbalances. Second, the paces of revenue and spending decentralization are different. Overall, local governments’ spending share rose twice as much as revenue share during 1988–2006. The increase in local governments’ spending share was broadly steady during the entire period, while the rise in their revenue share accelerated after the renegotiation round in 1996.

The different paces of decentralization in revenue and spending have led to changing vertical fiscal imbalances. For an empirical measure, we follow the literature (see e.g., Eyraud and Lusinyan 2013) to define vertical fiscal imbalance (VFI) as the share of the local governments’ own spending (including debt interest payment) that is not financed by its own revenue:

\[
VFI = \frac{\text{Regional spending} - \text{Regional revenue}}{\text{Regional spending}} \times 100
\]

29 We choose to focus on patterns up until 2006 because the European debt crisis and Global Financial Crisis from 2007 onward likely disrupted the intergovernmental fiscal relations in many countries, while the sharp decline in economic activity and the bailout of the financial sector led to drastic increases in public debt in many countries that were largely unrelated to vertical fiscal arrangements. Among related works, Eyraud and Lusinyan (2013), Sorribas-Navarro (2011) also stop data in 2006 or 2007, as do many other papers that look at economic patterns and trends of European countries before the crisis, e.g., García-Santana et al. (2020). 

Panel 5c shows that the sluggish revenue decentralization from the mid-1980s to mid-1990s led to rising vertical fiscal imbalances. This was reversed post-1996 when revenue decentralization caught up and resulted in a general trend of falling vertical fiscal imbalances from mid-1990s to mid-2000s.

Eyraud and Lusinyan (2013) among others document a negative correlation between a country’s vertical fiscal imbalances and aggregate fiscal performance. Similarly, our model predicts a positive relationship between vertical fiscal imbalances and total government debt (Section 2.4.3). Consistent with these findings, Panel 5d shows that total government debt in Spain increased during 1988–1996 when the country’s vertical fiscal imbalances were rising, and fell post-1996 when vertical fiscal imbalances were falling.31

Regional debt and transfer. The transfer function (2) in our model implies a positive correlation between a region’s debt stock and its transfer income. Because of this correlation, central government’s transfer lowers local governments’ cost of borrowing and hence leads to overborrowing in the model. Using regional transfer income data from the Economic Database of the Spanish Public Sector (BADESPE) and regional debt data from Banco de España, we document such a positive relationship for Spain. The left panel of Figure 6 plots the lagged debt/GDP and gross transfer/GDP for each of the 15 regions (excluding Navarra and the Basque country) in each year over 1994–2006, where each circle represents a region-year observation.32 We use lagged debt to reduce the possibility of reverse causality. For a clearer cross-region pattern, we plot the averages of a region’s debt/GDP

---

31 The second period roughly coincided with the housing boom in Spain. Increased transactional tax income likely also contributed to the falling total government debt during that period. We do not model the housing boom, but by using local and central government revenue directly from the data, our quantitative exercise captures the revenue effect from the housing boom. In other words, increased local revenue from the housing boom affects the central-local fiscal arrangement in the same way as other local revenue: it narrows the vertical fiscal imbalances and reduces local governments’ reliance on central transfers.

32 We start the sample in 1994 due to availability of regional debt data. For historical reasons, Navarra and the Basque country (País Vasco) have a different fiscal regime from the other regions. We follow the literature and exclude them from cross-region comparisons.
Figure 6: Relationship between Spain’s regional debt and regional transfer income

Note: Left panel plots the region’s transfer income from central government (normalized by regional GDP) against its lagged regional debt-to-GDP ratio for each year. Each circle is a region-year observation. Right panel plots each region’s average transfer income from central government (normalized by regional GDP) against its average regional debt-to-GDP ratio during 1994–2006. Each circle represents one region.

and gross transfer/GDP over 1994–2006 in the right panel. Both plots show a clear positive correlation between a region’s debt and its intergovernmental transfer revenue. This positive correlation in Spain is consistent with the findings of other empirical studies. For example, Sorribas-Navarro (2011) conducts regression analysis to study the relationship between regions’ (lagged) debt and their intergovernmental grants received from the Spanish central government. Controlling for a range of economic (income, unemployment), size (population, migration) and political factors, she finds a positive and statistically significant relationship between a region’s indebtedness and grants received. Additionally, she also finds similar results from regressions using instruments to control for potential endogeneity of debt.  

To complement Sorribas-Navarro (2011)’s finding, in Appendix B.2 we use an approach following Rodden (2005) to show that the financial markets expect the Spanish central government to transfer more to the more indebted regions. We show that in Spain, the dispersion of S&P ratings across regions is very small, despite the sizable heterogeneity in regional debt burden (captured by debt-to-own revenue ratio). This is in contrast to countries like Canada where regional credit ratings are much more dispersed than Spain, which implies that federal bailout or debt-dependent transfers are generally not expected. The exercise shows that the financial markets (and hence by extension the other market participants, such as local governments) anticipate that the more indebted regions in Spain do not pose higher default risks, because they will receive more central-to-local transfers.

The endogeneity problem may arise if, for example, an exogenous negative shock to a region induces the region to increase its debt, while at the same time the central government distributes more transfers to this region due to existing insurance arrangement (e.g., federal unemployment insurance or disaster insurance payments). To take into account this potential endogeneity, Sorribas-Navarro (2011) uses the share of debt at a variable interest rate interacted with the change in interbank interest rate as the instrument. Note that the accumulation of debt comes from two parts: primary deficits and interest payment. The instrument is by construction correlated with interest payment through the share of variable-rate debt. But by using the interbank interest rate, which is a national rate, the instrument is arguably uncorrelated with other region-specific shocks.
to help them with debt service costs. In addition to the empirical evidence, there is also anecdotal evidence to support the positive relationship between debt and transfers in Spain and several other countries. The details are included in Appendix B.2.

3.2 Calibration

We calibrate the parameters by matching the Markov equilibrium steady state moments to the empirical counterparts in Spain. All debt, spending, revenue, and transfer moments are normalized by trend national GDP. To quantify the impact of decentralization on debt changes, we look at two separate episodes: the rising vertical fiscal imbalances and debt during 1988–1996, and the falling imbalances and debt during 1996–2006. Accordingly, we calibrate the model separately to the economies in 1988, 1996, and 2006.

We introduce two additional parameters to make the model more flexible to fit the data. First, we allow the constrained regions to borrow, subject to a debt limit $b^o \leq B^o$. When $B^o$ is small enough, which is the case in our calibration, the debt limit always binds and each constrained region’s debt is $B^o$ in equilibrium. These regions then borrow at the interest rate $R(B^o)$ in equilibrium. This modification allows us to better match the constrained regions’ debt level observed in the data. Second, to better match the unconstrained regions’ debt level, we introduce an exogenous parameter $\epsilon$ that adjusts the distortionary effect of debt-dependent transfer rate $\tau$. More details about $\epsilon$ will be given later. Appendix C provides the general model with these two additional parameters. The model presented in Section 2 is a special case of the general model with $B^o = 0$ and $\epsilon = 1$, which we use to illustrate the main mechanisms of the model. We emphasize that these two additional parameters do not change the theoretical results of Section 2. Appendix E shows the proofs of the theoretical results with $\epsilon \in (0, 1]$ and $B^o \geq 0$.

3.2.1 Exogenously set parameters

We set some parameters exogenously using data moments. Table 2 summarizes these externally calibrated parameters.

**Constrained regions’ share $(1 - \mu)$ and debt limit $(B^o)$.** The constrained regions in our model can be best represented by the regions with low and stable debt levels in the data. We rank all 15 Spanish regions (excluding Navarre and the Basque country) by their debt level and debt volatility. We use the 7 regions with the lowest average debt levels and debt volatility to proxy for the constrained regions. During 1994–2006, the shares of national GDP and population of these 7 regions were about 20%, which we use as the value for $1 - \mu$. The average local public debt of these 7 regions was stable over time at around 4% of their GDP, which we use for the constrained regions’ debt limit $B^o$.

---

34 The model is solved by approximating the equilibrium policy functions as functions of current states. Appendix H provides more details.

35 Appendix E gives the criterion for how small $B^o$ needs to be.

36 During 1994–2006, the average regional debt-regional GDP ratio for the constrained regions was 4%, compared to 11.5% for the unconstrained regions. The standard deviation of regional debt-regional GDP ratio during the same period had an average of 1.1 among the constrained regions, compared to 2.1 for the unconstrained regions.
Table 2: Externally Calibrated Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>Time discount factor</td>
<td>0.95</td>
</tr>
<tr>
<td>$\psi_d$</td>
<td>Central interest elasticity</td>
<td>0.03</td>
</tr>
<tr>
<td>$\psi_b$</td>
<td>Local interest elasticity</td>
<td>0.03</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Unconstrained region’s Pareto weight</td>
<td>0.80</td>
</tr>
<tr>
<td>$B^o$</td>
<td>Constrained region’s debt limit</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$e$</td>
<td>0.3117</td>
<td>0.3835</td>
<td>0.4041</td>
</tr>
<tr>
<td>$f$</td>
<td>0.0430</td>
<td>0.0638</td>
<td>0.1327</td>
</tr>
<tr>
<td>$\bar{T}$</td>
<td>0.032</td>
<td>0.07</td>
<td>0.0485</td>
</tr>
</tbody>
</table>

Note: Revenue and transfer statistics are normalized by national GDP.

In general, changing the values of $\mu$ and $B^o$ does not significantly affect the quantitative results once the other parameters of the model are also re-calibrated. As a robustness check, in Section 4.3.3, we present an alternate classification of constrained regions and the corresponding set of $\mu$ and $B^o$, which yield similar quantitative results.

**Time discount ($\beta$) and interest rate schedules ($R(\cdot), S(\cdot)$).** The annual time discount rate $\beta$ is set at 0.95. We use log utility functions. The exogenous interest rate schedules for local (unconstrained and constrained) and central debt are defined à la Schmitt-Grohe and Uribe (2003):

$$R(b) = 1/\beta + \psi_b(e^{(b - \bar{B})} - 1)$$

$$S(d) = 1/\beta + \psi_d(e^{(d - \bar{D})} - 1)$$

and satisfy Assumption 2. Grande et al. (2013) estimate that a 1-percentage-point increase in public debt-to-GDP ratio corresponds to a 3-basis-point increase in interest rates among OECD countries during 1995–2011. Based on this estimate, we set the elasticity parameter for central debt $\psi_d = 0.03$.

There is no consensus on the interest rate elasticity of local debt, and so in the baseline, we set $\psi_b$ equal to $\psi_d$. In Section 4.3.2, we use alternative values of $\psi_b$ and $\psi_d$ for sensitivity analysis.

**Revenue ($e, f$) and transfer ($\bar{T}$).** Data on government revenue, inter-governmental transfer, and central and local (regional) government debt come from Eyraud and Lusinyan (2013) and IMF’s Government Finance Statistics (GFS), and are normalized by trend national GDP. The lower bound $\bar{T}$ on uniform transfer is not directly observable in the data. Sorribas-Navarro (2011) documents that over 1987–1996, about 85% of central-to-regional transfers (grants) in Spain were non-discretionary (formula-based), and this ratio fell to 81% during 2002–2006. This non-discretionary transfer corresponds to the uniform transfer $T^u$ in our model. Because $T^u = \bar{T}$ (Proposition 1), $\bar{T}$ is then calculated as the product of the total central-to-regional transfer from the data and the share of non-discretionary transfers reported by Sorribas-Navarro (2011).\(^{37}\)

\(^{37}\)In the data, a government’s fiscal balances (flow) do not necessarily match the change in debt (stock). For example, the central government’s budget equation (in nominal terms) $\text{transfer} + \text{other non-interest expenditure} + \text{interest expense} = \text{revenue} + \text{change in debt}$ does not hold in the data. One reason for this statistical discrepancy is that we use gross debt instead of net debt to calculate change in debt, as is customary in the literature, due to the lack of accurate public asset data.
3.2.2 Endogenously determined parameters

Given the exogenously set parameters, we then jointly calibrate the remaining parameters.

Transfer negotiation. In our model, the central government can achieve larger redistribution between constrained and unconstrained regions by raising $\tau$. In reality, redistribution may face political challenges. A region may choose to stay out of a fiscal federation (autarky) if it feels the net intergovernmental transfers it receives is much lower than other regions. To ensure that all regions stay in the fiscal federation, the cross-region difference in transfer (or equivalently, the extent of cross-region redistribution) cannot be too big relative to the region’s cost of being in autarky.\footnote{This “participation constraints” problem is discussed, for example, in Persson and Tabellini (1996), who show that the threat of secession gives rise to a limit on the extent of regional redistribution. See also Balcells et al. (2015) and Porta (2015) on Catalonia’s complaints about the uneven inter-governmental transfers in Spain.} We introduce a negotiation term in the transfer function to capture this “participation constraint” in a reduced-form way. This change will allow us to better match the data moments of local and central debt, but will not affect the theoretical results presented in Section 2.

In particular, we replace the transfer function (2) with

$$T_{i,t} = \tau_t b_{i,t} + (1 - \epsilon) \tau_t (\bar{b}_t - b_{i,t}) + T_{i}^{u}$$

(35)

where $\bar{b}_t = \mu b_t^n + (1 - \mu) B^o$ is the cross-region average debt, and $\tau_t (b_t - b_{i,t})$ is the difference between the average and region $i$’s transfer incomes. Parameter $\epsilon \in [0,1]$ captures the negotiation power of central versus regions: a region gets $1 - \epsilon$ portion of the cross-region transfer difference. When $\epsilon$ is equal to 1, the central government has all the negotiating power, and the transfer rate $\tau$ is fully redistributive. Whenever $0 < \epsilon < 1$, $\tau$ is only partially redistributive. When $\epsilon = 0$, transfer has no redistributive effect at all. Accordingly, the unconstrained local government’s Euler equation becomes

$$u^n_g = \beta R(b^n')(1 - \epsilon \tau') u^{n'}_g$$

We let data inform us on the value of $\epsilon$ by jointly calibrating it with other parameters. In Section 4.3.4, we explore the quantitative effects of $\epsilon$.

Joint calibration. We have four parameters to be calibrated internally: $\bar{B}$ and $\bar{D}$ in the interest rate schedules, the central government’s share of spending responsibility $\theta$, and the negotiation parameter $\epsilon$ in the transfer function. We jointly calibrate them to match four steady-state moments from the model: (1) total local government debt, (2) central government debt, (3) central-local spending ratio, and (4) total central-to-local transfer. For the year 1988, we use the four moments to pin down the four parameters. For later years, we fix $\epsilon$ to the 1988 value and use the first three moments to calibrate the other three parameters.\footnote{In a previous version we allowed $\epsilon$ to change after 1988 and calibrated four parameters using all four moments in each of the three years, and found that the variation of calibrated $\epsilon$ is marginal in the three years.}

Intuitively, all else equal, a higher $\bar{B}$ or $\bar{D}$ raises debt levels; a larger $\theta$ raises the central-local
Table 3: Internally Calibrated Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Target</th>
<th>Calibrated Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dot{B}$</td>
<td>Local government debt</td>
<td>0.0128</td>
</tr>
<tr>
<td>$\dot{D}$</td>
<td>Central government debt</td>
<td>0.3</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Central-local spending ratio</td>
<td>0.789</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Central-to-local transfer</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Note: Debt and transfer levels are normalized by national GDP. *We assume $\epsilon$ for 1996 and 2006 is the same as 1988.

Table 4: Targeted and Untargeted Moments

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Targeted Moments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local government debt</td>
<td>0.062</td>
<td>0.062</td>
<td>0.130</td>
<td>0.130</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td>Central government debt</td>
<td>0.338</td>
<td>0.337</td>
<td>0.615</td>
<td>0.614</td>
<td>0.336</td>
<td>0.336</td>
</tr>
<tr>
<td>Central-local spending ratio</td>
<td>3.63</td>
<td>3.63</td>
<td>1.85</td>
<td>1.85</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>Central-to-local transfer</td>
<td>0.038</td>
<td>0.039</td>
<td>0.077</td>
<td>0.080*</td>
<td>0.056</td>
<td>0.055*</td>
</tr>
<tr>
<td><strong>Untargeted Moments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local spending</td>
<td>0.060</td>
<td>0.062</td>
<td>0.125</td>
<td>0.129</td>
<td>0.166</td>
<td>0.162</td>
</tr>
<tr>
<td>Central spending</td>
<td>0.217</td>
<td>0.224</td>
<td>0.231</td>
<td>0.239</td>
<td>0.174</td>
<td>0.170</td>
</tr>
<tr>
<td>Vertical fiscal imbalance</td>
<td>32.2</td>
<td>34.0</td>
<td>52.2</td>
<td>53.1</td>
<td>21.3</td>
<td>21.1</td>
</tr>
<tr>
<td>Local interest payment</td>
<td>0.0037</td>
<td>0.0032</td>
<td>0.0085</td>
<td>0.0066</td>
<td>0.0030</td>
<td>0.0062</td>
</tr>
<tr>
<td>Central interest payment</td>
<td>0.026</td>
<td>0.017</td>
<td>0.043</td>
<td>0.031</td>
<td>0.013</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Note: All variables except central-local spending ratio and vertical fiscal imbalance are normalized by national GDP. Out of the untargeted moments, there is one degree of freedom, given the targeted moments and the central and local governments’ budget constraints. Additionally for 1996 and 2006, because $\epsilon$ is assumed the same as 1988, there is one more degree of freedom (*).

spending ratio; a higher $\epsilon$ increases the distortion of transfers, raises total transfer, and leads to higher central and local government debt levels.\(^{40}\) Table 3 reports these internally calibrated parameters, and Table 4 presents the targeted and untargeted moments. We match the targeted moments well. In particular, the calibrated model captures three properties in the data: debt levels increased from 1988 to 1996 and fell from 1996 to 2006; total transfer also increased during the first period and fell during the second period; and central-local spending ratio fell during both periods. For the untargeted moments, the model also does reasonably well in matching untargeted moments in the calibrated years.

**Degree of freedom.** Out of the untargeted moments in Table 4, there is one degree of freedom,\(^{40}\)

\(^{40}\)We also tried an alternative calibration strategy where we set $\epsilon = 1$ and jointly calibrated the share of unconstrained regions $\mu$ (instead of taking it directly from data) with the other parameters. $\epsilon = 1$ means $\tau$ is fully redistributive and distortionary, while a smaller $\mu$ means that a smaller share of regions are subject to the distortionary effect of $\tau$, and so the aggregate distortion is limited. As such, we can loosely match data moments of total transfer and central debt using $\epsilon = 1$ and a very small $\mu$. However, we find that because of the large distortionary effect of $\tau$ when $\epsilon = 1$, unconstrained regions’ debt in the model is implausibly high—about 10 times higher than the level observed in the data.
after the four parameters \( \{\bar{B}, D, \epsilon, \theta\} \) are calibrated and the local and central governments’ budget constraints are taken into account. Additionally, for 1996 and 2006, because \( \epsilon \) is assumed the same as 1988, there is one more degree of freedom, i.e., two degrees of freedom in total.

**Out-of-sample model fit.** To further test the model fit, we use the model calibrated to 1988 to look at how the model performs in the *out-of-sample* years 1989–2006. More specifically, we fix all parameters at their 1988 values, and then only change the fiscal decentralization parameters \( \{\epsilon, f, \theta\} \) from their 1988 levels to the levels for each year during 1989–2006. \( \epsilon \) and \( f \) are taken directly from the data. Because \( \theta \) governs spending decentralization but is not directly observed in the data, we take its calibrated values in 1988 and 2006 and linearly interpolate for the out-of-sample years in between.

Figure 7 compares the total government debt, vertical fiscal imbalance (VFI), and local and central government spending moments from the model and the data during the out-of-sample years. For this exercise, since the model is not calibrated for the out-of-sample years of 1989–2006, all moments are in fact free moments. The out-of-sample prediction for total government debt presents a hump-shaped pattern observed in the data. All other key moments follow the data closely, which provides support for the model and the initial steady state calibration. Note that although the parameters \( B \) and \( D \) are important for the debt levels in the model, we keep both parameters unchanged at the 1988 values in this exercise. That is why in Figure 7 we can only match the general trend but not 100% of the changes in debt. We explore the difference in debt between the data and the model moments in Section 4.1, where we use counterfactual experiment to show how much fiscal decentralization can explain the change in debt.

**Figure 7: Out-of-Sample Model Fit**

![Graphs comparing data and model moments](image)

Note: Plot compares the data and model moments. The model is calibrated in 1988, and for the later years, we only change parameters related to revenue decentralization (\( \epsilon \) and \( f \) directly from the data) and spending decentralization (\( \theta \) linearly interpolated). All other parameters are fixed at their 1988 values.
This section uses the calibrated model for two applications. First, we quantify the effects of fiscal decentralization on public debt accumulation, under the calibrated Markov equilibrium where the central government cannot pre-commit to future transfer policies. Second, we quantitatively compare the Markov equilibrium to the Ramsey equilibrium and the allocation under a consolidated government. We provide extensions and sensitivity analyses at the end of the section.

4.1 Fiscal decentralization

As discussed in Section 2.4.3, the model predicts that a revenue decentralization alone narrows vertical fiscal imbalances, lowers transfers, and reduces government debt, whereas a spending decentralization widens vertical fiscal imbalances, raises transfers, and increases government debt. In this section, we perform quantitative experiments to see how much of the changes in government debt in Spain during 1988–2006 can be attributed to the unbalanced fiscal decentralization reforms. Figure 5 shows that the rise and fall in total government debt coincide with major changes in the decentralization reforms in Spain. For this reason, we divide 1988–2006 into two episodes: 1988–1996 characterized by faster spending than revenue decentralization and rising total government debt, and 1996–2006 with revenue catching up with spending decentralization and falling total government debt.

Table 5 summarizes the results. To construct the counterfactuals, we change only the parameters related to fiscal decentralization and keep all other parameters unchanged. More specifically, for the 1988–1996 period, we start with the model economy calibrated to data moments in 1988, and then change the government revenue and spending responsibility parameters \((e, f, \theta)\) to their 1996 values to get a “counterfactual 1996” economy. Comparing the actual change in debt during 1988–1996 (Column 1 to Column 2) with the counterfactual change (Column 1 to Column 3) tells us the effects of decentralization on debt. For example, the total government debt-to-GDP ratio increased from 0.399 to 0.744 in the data, and to 0.534 in the counterfactual experiment. This indicates that the unbalanced decentralization, in this case, the faster spending than revenue decentralization, generates 39% of the rise in total government debt during 1988–1996. Similarly, for the 1996–2006 period, we start with the model economy calibrated to data moments in 1996, and then change \((e, f, \theta)\) to their 2006 values to obtain the “counterfactual 2006” economy. The comparison between the actual change in debt during 1996–2006 (Column 4 to Column 5) and the counterfactual change (Column 4 to Column 6) tells us the contribution of fiscal decentralization to debt changes during the second period. Figure 8 shows the actual and counterfactual total government debt in each period and illustrates the change in debt due to fiscal decentralization.

Overall, the unbalanced revenue and spending decentralization explains 39% of the actual changes in total government debt from 1988 to 1996 and 18% from 1996 to 2006. The rest are explained by changes in uniform transfer and interest rate schedule parameters \(\bar{B}\) and \(\bar{D}\). Decentralization explains a smaller proportion of the change in debt during 1996–2006 than during 1988–1996. One
Figure 8: Illustration of Counterfactual Experiment

(a) 1988–1996 period
(b) 1996–2006 period

Note: Counterfactual experiments change decentralization parameters \((e, f, \text{ and } \theta)\) to the end year values, and keep all other parameters at their starting year values. The experiments illustrate the change in debt that is explained by decentralization.

Table 5: Counterfactual Experiment
Effects of fiscal decentralization on debt levels

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Local debt</td>
<td>0.062</td>
<td>0.130</td>
</tr>
<tr>
<td>Central debt</td>
<td>0.337</td>
<td>0.614</td>
</tr>
<tr>
<td>Total government debt</td>
<td>0.399</td>
<td>0.744</td>
</tr>
</tbody>
</table>

Note: Total government debt (also known as “general government debt”) is the sum of local and central government debt. Counterfactual experiments use revenue \((e)\) and spending responsibility share \((\theta)\) values from end year, and keep all other parameters at their starting year values. Numbers in parentheses show the percent of debt changes that is explained by decentralization. All debt levels are normalized by national GDP.

possible reason is that there was increased scrutiny over government borrowing during the later period in order to meet EU regulations, such as the Stability and Growth Pact. That made it harder for all levels of governments to borrow and could have weakened the effect of soft budget constraint.

For both episodes, fiscal decentralization explains a larger proportion of changes in local debt and a smaller portion for central debt. In an alternative calibration presented in section 4.3.2, we use a more elastic local interest rate schedule \((\psi_b > \psi_d)\), as motivated by anecdotal evidence. The more elastic local interest rate increases the cost of local relative to central borrowing. As such, fiscal decentralization explains larger changes in central government debt than in the baseline.

**Debt-neutral (balanced) fiscal decentralization.** Spending decentralization alone raises debt, so how much revenue decentralization is needed to make a decentralization reform debt-neutral? Figure 9 plots the relationship between revenue and spending decentralization such that the total gov-
Government debt remains unchanged. For the purpose of illustration, the graph uses 2006 (green cross) as the base year. Points on the black line represent economies with different decentralization levels \((f/e, 1 - \theta)\), but the same total government debt. The upward sloping line indicates that a spending decentralization requires a proportionate increase in local governments’ share of revenue for the reform to be debt neutral. Economies above the line (e.g., 1988) have lower government debt than 2006, because their vertical fiscal imbalances are smaller and local governments are less reliant on transfers compared to the 2006 economy. Economies below the line (e.g., 1996) have larger vertical fiscal imbalances relative to 2006, and as a result, local governments rely more heavily on transfers and total government debt levels are higher than the 2006 economy.

**Figure 9:** Debt-neutral fiscal decentralization

Note: An increase in the local share of spending responsibility \((1-\theta)\) corresponds to a spending decentralization. An increase in local revenue as a share of total revenue \((f/e)\) corresponds to a revenue decentralization. Solid black line plots debt-neutral fiscal decentralization from the 2006 economy (green cross). Each point on the line has different \((f, \theta)\) values but the same steady state total (sum of central and local) government debt as the 2006 economy. All parameters except for \(\theta\) and \(f\) are held at calibrated 2006 values. Blue circle marks the 1988 economy, and red dot marks the 1996 economy.

This exercise should not be interpreted as an argument against spending decentralization. For example, the model ignores the potential efficiency gain of spending decentralization due to local governments’ better knowledge about local needs. The emphasis here is that spending decentralization should go hand-in-hand with revenue decentralization to maintain a balance between local governments’ own revenue and their spending responsibilities. An unbalanced fiscal decentralization can increase the reliance of local governments on transfers, which leads to overborrowing and a deterioration in the aggregate fiscal performance.

### 4.2 Comparison with other fiscal systems

In this section, we compare the steady state allocations and welfare of three different fiscal systems: (1) a consolidated government, (2) two layers of governments, where the central government cannot pre-commit to future policies (Markov equilibrium), and (3) two layers of governments, where the central government can commit to future policies (Ramsey equilibrium). The comparison in
Table 6 provides a numerical illustration of the theoretical results summarized in Table 1, using the parameters calibrated to data moments in 1996.

Consistent with the theoretical results, under a consolidated government, the steady state unconstrained local government debt $b^n$ equals $\bar{B}$, and central debt $d$ equals $\bar{D}$. The two types of local governments have the same spending, $g^n = g^o$, and the adjusted ratio of local-central marginal utilities, $(1 - \theta) [\mu u^n_g + (1 - \mu) u^o_g] / (\theta v_c)$, is equal to 1, i.e., $\text{MULG} = \text{MUCG}$.

In the non-commitment economy (Markov equilibrium), the steady state total government debt is about 9% of GDP higher than the level under a consolidated government. Consistent with Lemma 1, unconstrained local governments have lower spending, $g^n < g^o$. The adjusted local-central marginal utility ratio is less than 1, i.e., $\text{MULG} < \text{MUCG}$, because the central government without commitment over-transfers relative to the level of transfers required for $\text{MULG} = \text{MUCG}$ (Proposition 3).

In the commitment (Ramsey) economy, because the central government has commitment to future policies, it internalizes the overborrowing incentives from positive $\tau$ and so it sets $\tau = 0$ at the steady state. Transfers in this economy are solely made through uniform transfer $T^u$ at the steady state, and $\text{MULG} > \text{MUCG}$ (Proposition 4). Without any overborrowing incentives ($\tau = 0$), both local and central debt levels are the same as in the consolidated government allocation (Corollary 1), and the cross-region spending gap is smaller than in the Markov equilibrium.

In terms of consumption-equivalent welfare, because of the larger debt and cross-region spending gap, the Markov economy has 0.98% lower welfare than the economy with a consolidated government. By contrast, because there is no overborrowing in the Ramsey economy, the welfare difference from the consolidated government allocation is much smaller (0.002%). Furthermore, a 5.1% prudential tax on local government debt, as described in Section 2.6, can help implement the consolidated government allocation.

4.3 Extension and robustness

4.3.1 Extension: internalizing interest rate schedules

In the baseline model, we assume that governments do not internalize the slopes of their interest rate schedules. The assumption gives us clean theoretical results, and allows us to differentiate from the literature on the pecuniary externality of decentralized borrowing.

In this section, we relax this assumption by allowing governments to internalize the slopes of their interest rate schedules and re-calibrate the model economy. Overall, relative to the baseline model, the only difference in the theoretical results is that the Ramsey steady state $\tau$ is no longer zero. All other theoretical results in the Markov equilibrium hold as before.\(^{41}\) Below we briefly describe the changes to the model and discuss the quantitative results.

**Model with extension.** When the consolidated government internalizes the slopes of the interest rate schedules, an extra term with the derivative of the interest rate schedules shows up in the opti-
Table 6: Comparison of Steady States

<table>
<thead>
<tr>
<th>Moments</th>
<th>Consolidated government allocation</th>
<th>Non-commitment (Markov)</th>
<th>Commitment (Ramsey)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform transfer, $T_u$</td>
<td>–</td>
<td>0.07</td>
<td>0.078</td>
</tr>
<tr>
<td>Distortionary transfer</td>
<td>–</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>Total central-local transfer</td>
<td>–</td>
<td>0.08</td>
<td>0.078</td>
</tr>
<tr>
<td>Unconstrained local debt, $b^n$</td>
<td>0.114</td>
<td>0.152</td>
<td>0.114</td>
</tr>
<tr>
<td>Central debt, $d$</td>
<td>0.585</td>
<td>0.614</td>
<td>0.585</td>
</tr>
<tr>
<td>Cross-region spending gap, $g^n - g^n$</td>
<td>0</td>
<td>0.0057</td>
<td>0.0038</td>
</tr>
<tr>
<td>Adjusted local-central marginal</td>
<td>1</td>
<td>0.98</td>
<td>1.00013</td>
</tr>
<tr>
<td>utility ratio, MULG/MUCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumption equivalent welfare change</td>
<td></td>
<td>–</td>
<td>-0.98</td>
</tr>
<tr>
<td>relative to consolidated govt allocation (%)</td>
<td></td>
<td></td>
<td>-0.002</td>
</tr>
</tbody>
</table>

Note: Parameters calibrated to the 1996 economy are used. Transfer, debt, and spending are normalized by national GDP. Consumption equivalent welfare change calculates the percent of local and central spending that the household is willing to forgo.

mality conditions (4)–(5):

\[
\beta v_c(c_{t+1}) = \frac{1}{S(d_{t+1})} - \frac{d_{t+1}S_d(d_{t+1})}{S(d_{t+1})^2}
\]

\[
\beta u_g(g^n_{t+1}) = \frac{1}{R(b^n_{t+1})} - \frac{b^n_{t+1}R_b(b^n_{t+1})}{R(b^n_{t+1})^2}
\]

The additional terms show up because the consolidated government internalizes that higher debt increases the marginal cost of borrowing. As such, optimal debt levels will be lower in this alternative setup.

In the economy without a consolidated government, each local government internalizes the interest rate cost of higher borrowing. The representative unconstrained region’s Euler equation also has an additional term with the derivative of its interest rate schedule

\[
u_g(g^n_t) \left[ 1 - \frac{b^n_{t+1}R_b(b^n_{t+1})}{R(b^n_{t+1})} \right] = \beta R(b^n_{t+1})(1 - \tau_{t+1}) u_g(g^n_t)
\]

The central government internalizes the interest rate schedules of both local and central debt, and as a result, there are additional terms with $S_d$ or $R_b$ in the central government’s optimality conditions.

Quantitative results. We re-calibrate the extended model and show that, similar to the baseline model, (1) there are sizable overborrowing and over-transfer in the Markov equilibrium, and (2) fiscal decentralization changes debt levels in quantitatively meaningful ways.

The results in Table 7 show that similar to the baseline model, the Markov equilibrium has higher local and central debt than the consolidated government allocation. Also similar to the baseline, the Markov government over-transfers (MULG < MUCG) and the Ramsey government under-transfers (MULG > MUCG). Two things differ qualitatively from the baseline. First, the debt levels un-
under the consolidated government are lower than (instead of equal to) the re-calibrated parameters \( B (=0.236) \) and \( D (=1.063) \), because the central and local governments internalize the change in interest rates from additional borrowing. Second, the Ramsey central government chooses a positive (instead of 0) distortionary transfer, and as a result, local debt is slightly higher than (instead of equal to) in the consolidated government allocation.

The quantitative effects of fiscal decentralization in the extended model are summarized in Appendix I.2. Overall, the results are similar with the baseline model: decentralization explains 32% of the increase in total government debt during 1988–1996 and 9% of the debt decrease during 1996–2006.

**Table 7**: Comparison of Steady States when governments internalize interest rate schedules

<table>
<thead>
<tr>
<th>Moments</th>
<th>Consolidated government allocation</th>
<th>Non-commitment (Markov)</th>
<th>Commitment (Ramsey)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform transfer, ( T^u )</td>
<td>( - )</td>
<td>0.07</td>
<td>0.0797</td>
</tr>
<tr>
<td>Distortionary transfer</td>
<td>( - )</td>
<td>0.0121</td>
<td>0.0001</td>
</tr>
<tr>
<td>Total central-local transfer</td>
<td>( - )</td>
<td>0.0821</td>
<td>0.0798</td>
</tr>
<tr>
<td>Unconstrained local debt, ( b^n )</td>
<td>0.1213</td>
<td>0.152</td>
<td>0.1216</td>
</tr>
<tr>
<td>Central debt, ( d )</td>
<td>0.593</td>
<td>0.615</td>
<td>0.593</td>
</tr>
<tr>
<td>Cross-region spending gap, ( g^o-g^n )</td>
<td>0</td>
<td>0.0053</td>
<td>0.0039</td>
</tr>
<tr>
<td>Adjusted local-central marginal utility ratio, MULG/MUCG</td>
<td>1</td>
<td>0.978</td>
<td>1.00015</td>
</tr>
<tr>
<td>Consumption equivalent welfare change relative to consolidated govt allocation (%)</td>
<td>( - )</td>
<td>-0.64</td>
<td>-0.0054</td>
</tr>
</tbody>
</table>

Note: Assuming governments internalize interest rate schedules. Parameters are re-calibrated such that the simulated Markov equilibrium moments match the targeted data moments in 1996. Transfer, debt, and spending are normalized by national GDP. Consumption equivalent welfare change calculates the percent of local and central spending that the household is willing to forgo.

### 4.3.2 Alternative calibration of interest rate elasticities \( \psi_b \) and \( \psi_d \)

In the baseline calibration, we assume that the elasticities of local and central debt interest rate schedules are the same: \( \psi_b = \psi_d = 0.03 \). However, it is also plausible that local governments face a more elastic interest rate schedule than the central government, as the market for local government debt is usually less liquid. In an alternative calibration, we let \( \psi_b = 0.05 \) and \( \psi_d = 0.01 \) and re-calibrate the model.\(^{42}\) As \( \psi_b \) becomes larger (and \( \psi_d \) becomes smaller), local (central) debt interest rates become more (less) sensitive to changes in local (central) debt. So local (central) debt will be less (more) responsive to fiscal decentralization than in the baseline calibration.

Table 8 shows that in terms of the changes in central and local debt, this alternative calibration performs better than the baseline calibration. For example, fiscal decentralization explains 23% of the increase in total government debt during 1988–1996 and 9% of the debt decrease during 1996–2006.

\(^{42}\) Appendix I.3 presents re-calibrated parameter values for each year.
(97%) of the actual change in central (local) debt during 1988–1996, compared with 17% (129%) in the baseline. Overall, this alternative calibration delivers similar changes in total government debt as the baseline: decentralization accounts for 37% of total government debt changes during 1988–1996, and 24% during 1996–2006.

### Table 8: Counterfactual Experiment
with alternative local and central debt interest elasticities

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>1988</td>
<td>1996</td>
<td>Counterfactual</td>
<td>1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1996</td>
<td>2006</td>
</tr>
<tr>
<td>Local debt</td>
<td>0.062</td>
<td>0.130</td>
<td>0.128 (97%)</td>
<td>0.130</td>
</tr>
<tr>
<td>Central debt</td>
<td>0.338</td>
<td>0.615</td>
<td>0.401 (23%)</td>
<td>0.615</td>
</tr>
<tr>
<td>Total government debt</td>
<td>0.400</td>
<td>0.745</td>
<td>0.529 (37%)</td>
<td>0.745</td>
</tr>
</tbody>
</table>

Note: Using $\psi_b = 0.05$ and $\psi_d = 0.01$. All other model parameters are re-calibrated. Calibration strategy is identical to the baseline. Total government debt (also known as “general government debt”) is the sum of local and central government debt. Counterfactual experiments use revenue ($e$ and $f$) and spending responsibility share ($\theta$) values from end year, and keep all other parameters at their starting year values. Numbers in parentheses show the percent of debt changes that is explained by decentralization. All debt levels are normalized by national GDP.

4.3.3 Alternative calibration of constrained regions’ share and debt limit: $1 - \mu$ and $B^o$

In the baseline calibration, we use half of the regions (7 out of 15) with the lowest debt level and volatility in the data to proxy for the constrained regions in the model. This gives the share of constrained regions $1 - \mu = 0.2$ and their debt limit $B^o = 0.04$. As an alternative calibration, we use 9 out of 15 regions with the lowest and most stable debt levels to proxy for the group of constrained regions. This yields $1 - \mu = 0.4$ and $B^o = 0.046$. With this set of $(\mu, B^o)$, we re-calibrate the model. We find that changing $\mu$ and $B^o$ has little effect on the results from the counterfactual experiment once parameters are re-calibrated to match the moments in each year. Appendix I.4 reports the full calibration and results. To summarize, decentralization explains 39% of the increase in total government debt from 1988 to 1996, and 15% of the decrease in total government debt from 1996 to 2006.

4.3.4 Effect of a larger negotiation parameter $\epsilon$

The jointly calibrated negotiation parameter $\epsilon$ has a sizeable impact on the Markov equilibrium debt levels. Figure 10 shows that holding other parameters at their calibrated values, both the central and total local debt levels increase with $\epsilon$. This is because as $\epsilon$ increases, $\tau$ becomes more redistributive, and so the central government is more willing to increase $\tau$ to subsidize unconstrained regions. Higher $\epsilon \tau$ encourages local governments to borrow more, and the central government also has to borrow more to help finance the higher total transfers.

A natural question is how $\epsilon$ affects the results from the counterfactual experiment. We perform the counterfactual exercises using a larger $\epsilon$ value (0.1) than the baseline (0.014), keeping all other
Figure 10: Relationship between $\epsilon$ parameter and debt levels

![Diagram showing the relationship between $\epsilon$ and debt levels.]

Note: All other parameters are kept at calibrated levels in 1996. $\epsilon = 0$ corresponds to the efficient allocation under a consolidated government. The blue line plots the total local debt-to-GDP ratio $\mu b^n + (1 - \mu) B^o$ at the steady state, and the red line plots the central debt-to-GDP ratio $d$.

 parameters the same as the baseline. The results are reported in Appendix I.5. With this larger $\epsilon$ value, the counterfactual exercises generate much larger changes in debt than the baseline: fiscal decentralization explains over 90% of the changes in total government debt during 1988–1996 and 1996–2006. This is because a larger $\epsilon$ increases the overborrowing effect of $\tau$, which amplifies the effect of fiscal decentralization on changes in debt. Note that with the larger $\epsilon$ value, the model-generated debt levels are much higher and not matched to the data.

4.3.5 Alternative approach along the transition path

In the baseline counterfactual exercise, we quantify the effect of fiscal decentralization by comparing the steady state debt levels under different vertical fiscal arrangements, characterized by $e, f$ and $\theta$. The implicit assumption is that the 1988, 1996, and 2006 economies have reached steady states after each fiscal decentralization. An alternative approach is to relax the steady state assumption, and measure the debt changes at a point on the transition path between steady states. But one difficulty with this alternative is that we do not know the agents’ expected future path of fiscal decentralization, which in turn affects the impact of fiscal decentralization on the transition path. For example, to quantify the impact of fiscal decentralization on the debt accumulation during 1988–1996, the expected path of revenue and expenditure decentralization after 1996 also matters. However, it would be difficult to know whether the observed acceleration in revenue decentralization after 1996 was well anticipated during 1988–1996.

To illustrate that the quantitative results are robust in the alternative approach, we take one possible scenario for the expected path of fiscal decentralization. We assume that fiscal decentralization—characterized by changes to $e, f$, and $\theta$ in our model—took effect right after 1988, and was expected to stay constant afterward. Specifically, we start from the calibrated 1988 steady state economy, apply new values of $\{e, f, \theta, B, D, \epsilon\}$ and simulate the model until 1996. Note that the simulated path does not necessarily reach a steady state by 1996. We calibrate the new values of these parameters
such that the simulated moments in 1996 match the data moments in 1996. To give more details, the new government revenues \( \{e, f\} \) can be directly taken from their 1996 values. The new spending responsibility parameter \( \theta \) and other parameters \( \{\bar{B}, \bar{D}, \epsilon\} \) are not directly observed from the data, so we jointly calibrate them to target the total local government debt, central government debt, central-local spending ratio, and central-to-local transfer in the simulated 1996 economy along the transition path. Table I.6 reports the calibrated parameter values and moments.

We then conduct the counterfactual exercise. Calibration in this alternative approach is complicated because it entails calibrating to a point on the simulated transition path. For this reason, we only simulate until 1996 and do the 1988–1996 counterfactual as an illustration. For this counterfactual exercise, we keep all the parameters at their 1988 levels except for the fiscal decentralization changes to \( e, f, \) and \( \theta \). Table I.6 compares the calibrated and counterfactual moments, and Figure I.1 plots the simulated path of total debt and compares it to the counterfactual path. Overall, the unbalanced fiscal decentralization explains 18% of the increase in total debt during 1988-1996, somewhat smaller than the results using our baseline calibration.

5 Conclusion

This paper develops a dynamic infinite-horizon model of fiscal federation with vertical fiscal imbalances. Our model captures the main ingredients of the typical soft budget constraint problem: local governments have the autonomy to decide their spending and borrowing; and central-to-local transfers serve as a “common pool” for local governments. Compared to literature, our main contributions are the expansion of the analysis to infinite horizons and the incorporation of central government debt. From the theoretical perspective, these changes allow us to discuss the interaction between central government’s policy today and tomorrow, in addition to the interaction between central and local governments. From the quantitative perspective, these changes allow us to quantify the impact of fiscal decentralization reforms on both local and central government debt. Central government debt is often omitted in this literature due to the balanced-budget assumption, even though its change is empirically more relevant than local government debt.

In the model, when the central government cannot pre-commit the amount of future transfers, local governments have overborrowing incentives because they expect the central government to transfer more to regions that are more indebted—a common result in the literature. The more novel result from our infinite-horizon model is that the central government over-transfers to local governments, to the extent that residents’ marginal utilities from local public consumption is lower than from central public consumption. We apply the framework to show that, consistent with the empirical patterns documented in the literature, when a fiscal decentralization reform increases vertical fiscal imbalances, local governments become more reliant on transfers and total government debt rises. Quantitatively, we find that fiscal decentralization in Spain explains 39% of the increase in total government debt during 1988–1996 and 18% of the decrease during 1996–2006.

The experience of Spain is not unique. Fiscal decentralization is often unbalanced. In many coun-
tries, the decentralization in spending is faster than revenue, for economic or political reasons. This gives rise to large local fiscal gaps to be filled by intergovernmental transfers, which are often associated with common pool and soft budget constraint problems. Notable examples include the Bremen and Saarland’s long-term reliance on transfers following the initial bailout in Germany, and more recently, the rising local government debt and the debt swap program in China, just to name a few. The policy lesson we draw here is that, to avoid a deterioration in aggregate fiscal performance, a country should implement balanced fiscal decentralization, such that any additional spending responsibilities assigned to the local level come with a sufficient increase in revenue allocated to the local governments.

REFERENCES


---

43Economically, public finance literature following Musgrave (1959) argues that centralized tax collection can minimize inefficient tax competition among regional governments, while Tiebout (1956) and Oates (1972) advocate for decentralized spending to cater to different tastes of local residents. Politically, a central government may have strong incentives to hold a large share of revenue to weaken the political power of local governments. For example, Robinson and Torvik (2009) argue that soft budget constraint is politically rational because it increases the probability of politicians’ political survival.


Appendix for
“Decentralization and Overborrowing in a Fiscal Federation”

Si Guo, Yun Pei, and Zoe Xie

Table of Contents

A Empirical evidence on VFI and debt ........................................ 3

B Model assumptions: Remarks ..................................................... 5
  B.1 Micro-foundation of $\theta$ in a special case ......................... 5
  B.2 Motivation of transfer function (2) .................................... 6
  B.3 Exogenous interest rate schedules ...................................... 11

C General model ........................................................................ 13

D Derivations ............................................................................. 15
  D.1 Markov optimality conditions and GEEs ......................... 15
  D.2 Ramsey optimality conditions ......................................... 20

E Proofs ..................................................................................... 22
  E.1 Proof of Lemma 1 (Cross-region spending gap in Markov) ........ 22
  E.2 Proof of Proposition 3 (Over-transfer in Markov) .............. 23
  E.3 Proof of Proposition 2 (Minimum uniform transfer in Markov) 24
  E.4 Proof of Proposition 4 (Under-transfer in Ramsey) .......... 25

F Model extensions .................................................................... 27
  F.1 Allowing both types of local governments to borrow ........... 27
  F.2 Adding adjustment cost on $T^u$ ........................................ 33

G A three-period model .............................................................. 36

H Computational appendix ......................................................... 39

I Additional quantitative results .................................................. 40
  I.1 Additional moments for Section 4.1 ................................. 40

---

44GUO: International Monetary Fund (sguo@imf.org). PEI: University at Buffalo, SUNY (yunpei@buffalo.edu). XIE: World Bank; Federal Reserve Bank of Atlanta (xiexx196@gmail.com). The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, IMF management, the Federal Reserve System, the Federal Reserve Banks, the IBRD/World Bank or its executive directors.
I.2  Quantitative results for Section 4.3.1 ........................................ 41
I.3  Quantitative results for Section 4.3.2 ........................................ 42
I.4  Quantitative results for Section 4.3.3 ........................................ 43
I.5  Quantitative results for Section 4.3.4 ........................................ 44
I.6  Quantitative results for Section 4.3.5 ........................................ 45
**A  Empirical evidence on VFI and debt**

Figure 1 Panel B illustrates the relationship between a country’s vertical fiscal imbalances (VFI) and changes in total government debt over time, for Spain, Germany, and Portugal. Figure A.1 in this appendix supplements those figures and shows the same general pattern among 25 other OECD countries.

In addition, we run a fixed effect panel regression on these 28 countries to establish positive correlation between VFI and change in total government debt. We follow the regression specification in Eyraud and Lusinyan (2013), but instead of primary deficit, we use change in total government debt-to-GDP ratio as the dependent variable:

\[ \Delta d_{it} = \beta \times VFI_{it} + \gamma \times X_{it} + \alpha_i + \theta_t + \epsilon_{it} \]

where \( \Delta d_{it} \) is the change in country \( i \)'s debt-to-GDP ratio between years \( t - 1 \) and \( t \); VFI\(_t\) is the country’s VFI in year \( t \); \( X_{it} \) is a set of control variables, which include expenditure decentralization, lag of debt-to-GDP ratio, lag of output gap, a measure of trade openness, population, and rule of law; \( \alpha_i \) are country fixed effects, \( \theta_t \) are time dummies, and \( \epsilon_{it} \) is the error term. We follow Eyraud and Lusinyan (2013) and Aldasoro and Seiferling (2014) in deciding the control variables. Both papers empirically study the relationship between a country’s VFI and primary deficit or change of government debt. Table A.1 shows that a 10 percentage point increase in VFI is associated with a 1.2 to 1.3 percentage point acceleration in the growth of total government debt as percentage of GDP.

*Table A.1: Dependent Variable: Change of total government debt (% GDP)*

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFI</td>
<td>0.12**</td>
<td>0.13***</td>
<td>0.13***</td>
<td>0.12**</td>
</tr>
<tr>
<td>Expenditure decentralization</td>
<td>-0.32***</td>
<td>-0.38***</td>
<td>-0.31***</td>
<td></td>
</tr>
<tr>
<td>Lag of debt-to-GDP ratio</td>
<td>-0.11***</td>
<td>-0.13***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag of output gap</td>
<td>-0.20</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness (trade as % GDP)</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population (million)</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule of law</td>
<td>-4.5***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>356</td>
<td>356</td>
<td>327</td>
<td>273</td>
</tr>
<tr>
<td>Countries</td>
<td>28</td>
<td>28</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Overall R(^2)</td>
<td>0.40</td>
<td>0.45</td>
<td>0.52</td>
<td>0.51</td>
</tr>
<tr>
<td>Within R(^2)</td>
<td>0.23</td>
<td>0.29</td>
<td>0.38</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Note: Fixed effects estimation with robust standard errors clustered at the country level. *, **, *** for significance at the 10, 5, and 1 percent level, respectively. Sample restricted to 1985–2007 if available.
Figure A.1: Vertical fiscal imbalances (blue dash, left) vs annual changes in total government debt (red dotted, right)

Note: Vertical fiscal imbalances = (Local governments’ spending – Local governments’ own revenue)/Local governments’ spending × 100. Total government debt is the sum of central and local government debts. Annual changes in total government debt are calculated as the first difference of total government debt as percent of national GDP. For Belgium, Denmark, Israel, Italy, Luxembourg, and Netherlands, the debt data are not available for the early periods.
B Model assumptions: Remarks

B.1 Micro-foundation of $\theta$ in a special case

While all the results in Section 2 work for any utility functions that are smooth, increasing and concave (Assumption 1), the preference parameter $\theta$ has a micro-foundation with log utility functions and unitary elasticity of substitution among varieties of goods (Section 2.1). This section provides a derivation of this micro-foundation.

Assume there are infinitely many varieties of public goods, each indexed by $\omega \in [0, 1]$. Goods $\omega \in [0, \bar{\omega})$ are provided by the central government. Goods $\omega \in [\bar{\omega}, 1]$ are provided by the local governments.

The household’s preference over the basket of public goods is

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma}$$

where $\gamma > 0$ is the household’s risk aversion. The basket of public goods is given by $C = \left[ \int_0^1 q(\omega) \frac{\sigma-1}{\sigma} d\omega \right]^{\frac{\sigma}{\sigma-1}}$, where $\sigma > 0$ is the elasticity of substitution and $q(\omega)$ denotes the public spending on good $\omega$. Let $c$ and $g$ denote the central and local governments’ total spending on public goods respectively. Because all types of goods $\omega \in [0, 1]$ have the same weight in the consumption basket, each central and local governments optimally chooses to evenly distribute its total spending on each type of goods.

$$C = \left[ \int_0^{\bar{\omega}} q(\omega) \frac{\sigma-1}{\sigma} d\omega + \int_{\bar{\omega}}^1 q(\omega) \frac{\sigma-1}{\sigma} d\omega \right]^{\frac{\sigma}{\sigma-1}}$$

$$= \left[ \int_0^{\bar{\omega}} \left( \frac{c}{\bar{\omega}} \right) \frac{\sigma-1}{\sigma} d\omega + \int_{\bar{\omega}}^1 \left( \frac{g}{1-\bar{\omega}} \right) \frac{\sigma-1}{\sigma} d\omega \right]^{\frac{\sigma}{\sigma-1}}$$

$$= \left[ \bar{\omega} \left( \frac{c}{\bar{\omega}} \right) \frac{\sigma-1}{\sigma} + (1-\bar{\omega}) \left( \frac{g}{1-\bar{\omega}} \right) \frac{\sigma-1}{\sigma} \right]^{\frac{\sigma}{\sigma-1}}$$

Therefore,

$$U(C) = \left[ \bar{\omega} \frac{1}{\bar{\omega}} + (1-\bar{\omega}) \frac{1}{\bar{\omega}} \right]^{\frac{(1-\gamma)}{\sigma-1}} \cdot \left[ (1-\theta)g^{1-\frac{1}{\sigma}} + \theta c^{1-\frac{1}{\sigma}} \right]^{\frac{(1-\gamma)}{\sigma}}$$

where $\theta = \frac{\bar{\omega}^{1/\sigma}}{\bar{\omega}^{1/\sigma}+(1-\bar{\omega})^{1/\sigma}}$. In the special case when $\sigma = \gamma = 1$, we have

$$U(C) = (1-\theta) \log(g) + \theta \log(c) - A(\bar{\omega})$$

where $A(\bar{\omega}) = \bar{\omega} \log(\bar{\omega}) + (1-\bar{\omega}) \log(1-\bar{\omega})$. This is essentially equivalent to utility function (1) by assuming both $u(\cdot)$ and $v(\cdot)$ are log functions. The parameter $\theta$ can be interpreted as the utility weight of central public goods.
B.2 Motivation of transfer function (2)

The transfer function (2) in our model assumes that at least part of the central government transfers is dependent on regions’ debt burden. In this section, we provide some justifications for this assumption. We first provide anecdotal country examples where transfers are directly dependent on debt burden. We then summarize the empirical evidence on the relationship between regional government debt burden and intergovernmental transfers in Spain. Finally, we show that with some simplifying assumptions, the “debt-dependent transfer” feature in function (2) can be mathematically derived from other (more complicated) transfer schemes with greater real world relevance.

B.2.1 Anecdotal country cases

We provide a brief review of country examples where federal transfers reward the less fiscally disciplined regions that have higher debt burden or higher deficits. The purpose of this review is to show that in many countries, regional debt burden is indeed a factor when the central government decides the amount of transfers. We include past “deficits” in addition to “debt” in our discussion because debt is the accumulation of deficits over time.

The anecdotal examples discussed here broadly fall into two categories. Some of these examples present evidence that the central government favors the local governments that are more indebted (or have been running larger deficits) through higher transfers in “normal” times. Other examples are interpreted as “bailout”, a term that is often linked to emergency transfers during “crisis” times. Although the debt-dependent transfers in our model occur in every period, as we will see, the distinction between bailouts in crisis times and the debt-dependent transfers in normal times is often blurred in real world.

One of the most relevant examples is the “deficit grant” program in Hungary. The initial intention of the program was to help local governments to overcome the deficits accrued through “no fault of their own” and avoid local government bankruptcy. The grants were only made to assist local governments for mandatory spending. However, judging whether a local government’s deficits were accumulated through exogenous shocks or its lack of fiscal discipline can be difficult in practice. As documented in Wetzel and Papp (2003), the deficit grants softened the budget constraints of local governments, because local governments “can increase their grant revenues through behavior changes,” such as raising less revenue and/or increasing expenditures on mandatory services, with the expectation that “deficit grants may potentially fill any gap.” According to Wetzel and Papp (2003), about one third of the local governments were receiving deficit grants in 1999. A similar program in India is documented by McCarten (2003).

Another example is the intergovernmental transfers in the public health sector in Italy. Bordignon and Turati (2009) document that prior to the reform in the mid-1990s, the central government systematically used additional transfers to cover local governments’ debt accumulated from excessive health spending in previous years. On average, about 50% of the health deficits accumulated by local governments during 1995-1999 were later offset by additional central government transfers by 2002. Although the term “bailout” is used to describe these transfers, they do not necessarily happen under emergency, as commented by Bordignon and Turati (2009): “Bailing out of regional health deficits is not an exceptional occurrence in Italy. With an average delay of 2-3 years, the Central government usually ‘finds out’ that it has made a ‘mistake’ in computing the health needs of regions, and covers (part of) the past health deficits of regions by issuing government bonds and by taking care of Local Health Units debts.”

45Similarly, Pedreja-Chaparro et al. (2006) document that the central government in Sweden bailed out local governments 1697 times during 1974-1992. This high frequency makes it hard to justify that all bailouts happened under emer-
The additional federal transfers to the two indebted German regions (Lander) – Bremen and Saarland – provide another example where federal transfers are de facto dependent on local governments’ debt burdens. As noted in Rodden (2005b), Bremen and Saarland avoided expenditure adjustments and accumulated large amount of public debt in the 1980s, betting higher federal transfers to pay for their regional public debt. The Federal Constitutional Court’s decision in 1992 confirmed that the federal government has the obligation to make additional transfers to these two regions to help relieve the regional fiscal burden and maintain the norm of public service. Following the court decision, additional federal supplementary transfers were provided to Bremen and Saarland annually.

There are two points to be emphasized in Germany’s case. First, the additional transfers to Bremen and Saarland are “debt-dependent”: the court decision in 1992 was explicit that the debt burden is an important factor in deciding whether additional federal transfers are justified. As noted in Von Hagen et al. (2000), the court decision was based on the argument that a state with “extreme budgetary hardship” is entitled to receive additional transfers, and “the decision clarified an emergency exists in a Lander if the deficit/expenditure ratio and interest payment/tax ratios are more than twice the Land average” (Rodden 2005b). Second, in the case of Bremen and Saarland, the distinction between crisis time bailout and debt-dependent transfers in normal times is also blurred. Even prior to the court decision in 1992, (smaller) annual grants had been provided to these two regions since 1987 to mitigate the debt burden (Von Hagen et al. 2000). After the court decision in 1992, even though the additional supplementary transfers from the federal government were initially planned for 5 years, they were extended several times and the payment continued till 2005 (Stehn and Fedelino 2009). Arguably, the long time span of the supplementary transfers makes it less convincing to argue the transfers should be solely interpreted as an emergency bailout.

In the United States, the type of debt-dependent transfers analogous to our model is the interest ratesubsidy provided by the federal government. In the United States, the federal interest rate subsidy is provided in two ways: the indirect subsidy through the income tax exemption of municipal bond interest payment; and the direct interest rate subsidy on municipal bond issuance, such as the Build America Bond (BAB) program (Luby et al. 2021). Although the indirect subsidy through tax exemption is less discretionary (because the effective subsidy rates are determined by the bond purchasers’ federal income tax rates rather than the federal government’s willingness to subsidize), the BAB program is clearly a discretionary program aiming to help local governments secure infrastructure financing. Similar interest subsidy programs were also established in other countries by extended bodies of the federal government, such as municipal banks, development banks and municipal funds (see Peterson 2000 for a review). These development financing entities are usually owned by the central government, hence the subsidies they provide are eventually borne by the central government budget.

B.2.2 Debt-dependent transfers in Spain

A. Anecdotal evidence

In Spain, although there is no comprehensive documentation of a formal program similar to the “deficit grants” in Hungary or Italy, Garcia-Milà et al. (2002) document that “for some regions and types of expenditures additional grants were given to cover the historical debt that had been accumulated... this is a dollar-for-dollar increase in grants in response to borrowing, at least for part of the debt.” This should not be surprising, given the central-local fiscal relationship in Spain shares many similarities with Italy, especially in the area of health spending.
There were also at least two waves of explicit bailouts of local governments in Spain. The first wave was in 1983 when the central government assumed part of the local government debt, as documented in Pedreja-Chaparro et al. (2006). The second wave took place in 2012. In response to the crisis and mounting regional government debt, two regional liquidity funds were set up by the central government to provide cheaper loans to regional governments: the Fund for Financing Payments to Suppliers (“FFPP” in Spanish) that lent to local governments to pay for arrears owed to local suppliers; and the Autonomous Liquidity Fund (“FLA” in Spanish) designed to help with regional governments’ difficulties in debt refinancing. In essence, local governments that were facing surging interest rates or even losing access to the financial market can borrow cheap loans from the FFPP and FLA funded by the central government.\(^{46}\)

B. Empirical evidence

Empirical studies on the causal link between central government transfer and local governments’ debt burden mainly follow two approaches. The first approach uses econometric methods to estimate or test whether higher debt burden (or higher deficits in the past) indeed results in higher transfers. Solé-Ollé and Sorribas-Navarro (2012) use the municipal level panel data in Catalonia region during 1988-2006 to study how municipal governments respond to shocks to revenue, expenditure and grants. They find that a positive expenditure shock to municipal government is followed by an increase in grants from upper level governments. As stated by the authors, these grants are usually discretionary grants, and they “might have generated a moral-hazard problem, with municipalities having the incentive to undertake excessive expansion since additional capital spending will always be financed (at least in the long run) by higher tiers of government.” Sorribas-Navarro (2011) uses regional fiscal data in Spain during 1986-2006 to document that regions with larger debt burden tend to receive more federal transfers. The potential endogeneity problem is addressed by (1) adding cyclical economic indicators as control variables, such as unemployment, GDP, and population net flows; and (2) using the share of debt at a variable interest rate interacted with changes in interbank interest rates as an instrumental variable.\(^{47}\)

The second approach follows Rodden (2005a) to see if the financial market distinguishes regional governments based on their debt service abilities. If regions’ debt service abilities do not affect their debt ratings, it implies that the financial market believes that the central government will eventually help indebted regional governments with their debt burden, either through emergency bailout or periodic transfers in normal times. Rodden (2005a) compares the dispersion of regional debt ratings in countries where regional governments face “harder” budget constraints (Canada and the U.S.) and “softer” constraints (Germany and Spain) in the late 1990s. He finds that in Canada and the U.S., regional debt ratings are more dispersed, and the cross-region rating difference can be largely explained by the heterogeneity in regional debt burden (proxied by regional debt-to-own revenue ratio). To the contrary, the debt ratings of German Landers and Spanish regions are much less dispersed and closely follow their sovereign ratings, and debt-to-own revenue ratio is a weak factor

\(^{46}\)There are two reasons why the FFPP and FLA can provide cheap loans to regions. The first is that the FFPP and FLA can borrow from banks under the central government’s guarantee, which lowered the funding cost of the FFPP and FLA. The second is that the central government also made direct budgetary contributions to the two funds. For example, the initial budgetary contribution to the FFPP announced in 2012 was €6 billion (Minguez et al. 2012).

\(^{47}\)The main concern about endogeneity is that a negative shock to a region (e.g., flood in a region) can increase the region’s debt (due to higher expenditure) while trigger additional federal transfers through insurance channels. This may create an upward bias to the estimated impact of regional debt on federal transfers. The ideal instrumental variable should be correlated with debt burden but uncorrelated with other shocks that affect transfers. As discussed in Sorribas-Navarro (2011), the accumulation of debt comes from two parts: primary deficits and interest payment. The instrument adopted (share of debt at a variable interest rate interacted with interbank interest rates) would be correlated with interest payment (and hence debt burden), but plausibly uncorrelated with other regional shocks such as regional flood.
in explaining the (small) cross-region rating difference. These findings provide the evidence that the market believes that the more indebted regions in Germany and Spain will eventually receive more transfers from the central government. We replicate the exercise in Rodden (2005a) for Spain, using Canadian provinces as a comparison. We extend the sample to 2006, the year when the quantitative excise in Section 3 ends. Figure B.1 confirms the findings by Rodden (2005a) that compared to Canada, the credit ratings of Spanish regions were much less dispersed and less correlated with their debt burden.

B.2.3 Connection between transfer function (2) and cost-based transfer schemes

We show that the “debt-dependent” part of the transfer function (2) can be mathematically derived from the “cost-based” transfer scheme in real world.

In many countries, the amount of central-to-local transfer is determined by the so-called “cost method” (see Bahl 2000 for an overview). That is, in principle, the cost for each local government to provide public services above a minimum national standard is first estimated. The amount of intergovernmental transfers to each local government is then determined to fill the gap between the estimated cost for public service provision and local governments’ own revenues. Though some of the factors affecting the cost of public services are observable (e.g., population) and can be incorporated into the formula of calculating the amount of transfers, accurately estimating the cost of local public service provision is still difficult. Therefore, local governments’ historical expenditure data has a large impact on the negotiating process between the central and local governments, even though historical expenditure is not necessarily an unbiased indicator of the “real” cost of local public services.\footnote{One related example noted by López-Casasnovas and Rosselló-Villalonga (2014) is that in Spain, before the central government shifted its health service responsibility to regional governments, it intentionally lowered the budget expenditure for health spending in the last two years right before the shift. This tacit move by the central government gave it some advantages in the negotiation with local governments about the amount of transfers, as it lowered the basis for assessing the cost of health services in the years immediately after the devolution of health service responsibility. In later years, the advantage arising from asymmetric information (on assessing the cost of public services) gradually tilted towards local governments as they became familiar with the operations of these services.}

To describe this process, assume that the total transfers to a local government $i$ include a uniform part and a discretionary part.

\[ T_{i,t} = T^u_{i,t} + T^d_{i,t} \]  

(B.1)

All regions receive the same amount of uniform transfers $T^u_{i,t}$, but potentially different amount of discretionary transfers $T^d_{i,t}$. The determination of discretionary transfers $T^d_{i,t}$ that the local government $i$ receives can be
The first term on the right-hand side of (B.4) shows that transfers are partially dependent on debt outstanding.

Equation (B.3) implies that for local government to finance higher spending, it can be transformed to

\[ C(g_{i,t-1}, r_{i,t-1} b_{i,t-1}) = (f_{i,t-1} + T_{i,t-1}^d + T_{i,t-1}^u) \]

where \( T_{i,t-1}^u \) and \( T_{i,t-1}^d \) are the uniform and discretionary transfers at \( t - 1 \), \( \alpha \in [0,1] \) is a constant, \( f_{i,t-1} \) is the regional revenue, and \( r_{i,t-1} b_{i,t-1} \) is the net interest payment on debt, where net interest rate \( r \) is related to the gross interest rate \( R \) in the paper by \( r = 1 - 1/R \).

The term \( C(\cdot) \) is the central government’s estimated cost of local public services in period \( t - 1 \), which will be elaborated later. Thus, (B.2) means if the central government estimates that the local government’s total revenue is insufficient to cover the cost of providing public services at \( t - 1 \), it seeks to increase the transfers for period \( t \) by a factor of \( \alpha \), either as a result of the goodwill of the central government or the bargaining between the two levels of governments. Note that \( C(\cdot) \) takes into account not only the expenses on salaries and goods and services (captured by \( g_{i,t-1} \)), but also the interest payment of the debt associated with spending. The reason is that for some local expenditure items (especially capital spending items such as building a hospital or highway), upfront spending financed by debt is usually incurred. The interest cost associated with this type of capital spending should be counted as a cost of local public service.\(^{49} \)

As discussed, estimating the cost of public service is difficult. In practice, past expenditure is an important factor in assessing the cost of service. For simplicity, assume the central government’s estimate of the cost of local public service follows

\[ C(g_{i,t-1}, r_{i,t-1} b_{i,t-1}) = g_{i,t-1} + r_{i,t-1} b_{i,t-1} + \delta_{i,t-1} \]

where \( \delta_{i,t-1} \) is an i.i.d shock with zero mean.

Substituting the local government’s budget constraint, the discretionary transfer that local government \( i \) receives can be transformed to

\[
T_{i,t}^d = T_{i,t-1}^d + \alpha \left[ C(g_{i,t-1}, r_{i,t-1} b_{i,t-1}) - (f_{i,t-1} + T_{i,t-1}^d + T_{i,t-1}^u) \right] \\
= T_{i,t-1}^d + \alpha \left[ (g_{i,t-1} + r_{i,t-1} b_{i,t-1} + \delta_{i,t-1}) - (f_{i,t-1} + T_{i,t-1}^d + T_{i,t-1}^u) \right] \\
= T_{i,t-1}^d + \alpha \left[ f_{i,t-1} + b_{i,t} - b_{i,t-1} + T_{i,t-1}^d + T_{i,t-1}^u + \delta_{i,t-1} - (f_{i,t-1} + T_{i,t-1}^d + T_{i,t-1}^u) \right] \\
= T_{i,t-1}^d + \alpha (b_{i,t} - b_{i,t-1}) + \alpha \delta_{i,t-1}
\]

Equation (B.3) implies that for local government \( i \), increasing borrowing \( b_{i,t} \) (decided at \( t - 1 \)) by one dollar to finance higher spending \( g_{i,t-1} \) will result in an increase in discretionary transfer \( T_{i,t}^d \) by \( \alpha \) dollar.

More formally, aggregating (B.3) over \( t = 1, 2, ... \) and substituting into (B.1), we have

\[
T_{i,t} = T_{i,t}^d + T_{i,t}^u \\
= \alpha b_{i,t} - \alpha b_{i,0} + T_{i,0}^d + \sum_{k=0}^{t-1} \alpha \delta_{i,k} + T_{i,t}^u
\]

The first term on the right-hand side of (B.4) shows that transfers are partially dependent on debt outstanding.

\(^{49}\)For example, in the United States, the criteria for state and local governments to apply for federal cost reimbursement grants is guided by the Office of Management and Budget (OMB) Circular A-87 and the associated guidance, which states that certain types of interest cost incurred by state and local governments should be allowable for the purpose of federal cost reimbursement.

10
B.3 Exogenous interest rate schedules

The theoretical results in the paper rely on the exogenous interest rate schedules \( R(\cdot) \) and \( S(\cdot) \) being increasing in debt outstanding.\(^{50}\) This assumption is widely used in international finance literature following Schmitt-Grohe and Uribe (2003) to ensure that the steady state is not dependent on initial conditions. The exponential functional form of \( R(\cdot) \) and \( S(\cdot) \) is only used for numerical exercises and is not a necessary assumption for the model properties to hold. Below we offer two ways to motivate the assumption that interest rates are increasing in debt. In the first approach, the increasing interest rate schedule comes from investors’ preference for portfolio diversification. In the second approach, the increasing interest rate schedule comes from financial intermediation frictions. Despite the different mechanisms, both approaches yield increasing interest rate schedules that are consistent with our model assumption.

The first way to motivate the positive slopes of interest rate schedules follows Chang et al. (2015), where international investors favor international portfolio diversification. Assume there is a continuum of international investors who can invest in Spanish government bonds \( D_t \) and global bonds \( D^*_t \).\(^{51}\) The international investors’ problem is:

\[
C^*_t \max_{C^*_t, D^*_{t+1}, D_{t+1}} \sum_{t=0}^{\infty} \beta^t U(C^*_t)
\]

subject to the international investors’ budget constraint,

\[
C^*_t + (D_{t+1} + D^*_t) \left[ 1 + \frac{1}{\beta} \left( \frac{D^*_t}{D^*_t + D^*_t} - \bar{\psi} \right) \right] \leq Y^*_t + S^*_t D_t + S^*_t D^*_t
\]

In this problem, \( C^*_t \) and \( Y^*_t \) are the international investors’ consumption and income, \( D_t \) and \( D^*_t \) are their holdings of Spanish and global bonds, and \( S^*_t \) and \( D^*_t \) are the respective gross interest rates. The parameter \( \bar{\psi} \) denotes the steady-state share of Spanish bonds in international investors’ portfolio. We assume \( \bar{\psi} \in (0,1) \) so that investors prefer some degree of portfolio diversification. The parameter \( \Omega > 0 \) controls the size of the adjustment cost deviating from this steady-state portfolio allocation.

Denote \( \eta_t \) as the multiplier to the budget constraint and \( \psi_t = \frac{D_t}{D_{t+1}} \), the optimality conditions are:

\[
\beta^t U'(C^*_t) = \eta_t
\]

\[
\Omega(\psi_{t+1} - \bar{\psi}) = \frac{\eta_{t+1}}{\eta_t}(S^*_t + S^*_t - S^*_t)
\]

Along the balanced growth path, \( \frac{m_{t+1}}{m_t} \to \beta \) and the world interest rate \( S^*_t = \frac{1}{\beta} \). Hence, the Spanish bonds’ interest rate is

\[
S_{t+1} = \frac{1}{\beta} + \frac{\Omega}{\beta} \left( \frac{D_{t+1}}{D_{t+1} + D^*_t} - \bar{\psi} \right)
\]

which is an increasing function of \( D_{t+1} \).

The second way to motivate the positive slopes of interest rate schedules is through financial intermediation frictions. Below we follow the setup of Gabaix and Maggiori (2015), though other ways to model financial frictions (e.g., Fanelli and Straub (forthcoming)) can also work.

\(^{50}\) Assumption 2 states that both \( R(\cdot) \) and \( S(\cdot) \) are increasing and convex. However, convexity is not a necessary condition for our theoretical results to hold. What is essential is that the interest rate schedules are increasing functions with ranges large enough to ensure an interior solution. Convexity is a sufficient (but not necessary) condition to ensure the ranges of the interest rate schedules are large enough.

\(^{51}\) The global bonds here can be understood as German government bonds or U.S. treasury bonds.
We assume that the government bond market is only accessible to a unit mass of “financiers” (e.g., banks and investment funds), reflecting the real world environment that the retail bond market either does not exist or exists in a small scale in many countries. Each financier has zero capital to start with and lives for only one period. In each period, a financier can draw deposits $q_t$ from households at interest rate $S_t^*$ and use the proceeds to invest in government bonds at interest rate $S_t$. Here $S_t$ will be endogenously determined. $S_t^*$ can be exogenously tied to global interest rates or central bank policy rates. At time $t$, a financier’s value can be denoted as

$$V_t = \max_{q_t} (S_t - S_t^*) q_t$$

It is clear that if the financial market is frictionless and there is no restrictions on the range of $q_t$ that a financier can choose from, the non-arbitrage condition ensures that $S_t = S_t^*$. 

The key assumption here is that there are some forms of financial frictions that can limit financiers’ capacity of intermediating deposits into investments in government bonds. When this happens, the non-arbitrage condition $S_t = S_t^*$ may not hold anymore. Following Gabaix and Maggiori (2015), the particular type of financial friction we consider here is that the financiers can divert some of the invested assets away. More specifically, for a financier that draws deposits $q$ from households and uses the proceeds to invest in government bonds, the financier has the option to divert $\Gamma q_t^2$ and walk away. Because households anticipate that a financier has the option to divert assets and walk away, any deposit contract between households and financiers has to ensure that the value of any financier is larger than the assets that can be diverted away,

$$V_t \geq \Gamma q_t^2$$  \hspace{1cm} (B.8)

Therefore, the financier will choose $q_t$ to maximize his value,

$$V_t = \max_{q_t} (S_t - S_t^*) q_t \quad \text{subject to the constraint (B.8)}$$

This problem yields the solution that each financier will choose

$$q_t = \frac{1}{\Gamma} (S_t - S_t^*)$$  \hspace{1cm} (B.9)

That is, because of the friction (B.8), each financier’s investment in government bonds will be limited and hence the spread between $S_t$ and $S_t^*$ may not necessarily be eliminated. Integrating over the unit mass of financiers, we have

$$D_t = \frac{1}{\Gamma} (S_t - S_t^*)$$  \hspace{1cm} (B.10)

Therefore, the interest rate of government bonds is an increasing function of bond supply. The premium (over the risk-free rate $S_t^*$) comes from the financial intermediation cost.

---

52Here we assume that the government bonds are one-period bonds sold at face value, so $q_t$ deposits can be used to purchase $q_t$ units of bonds. The principal and coupon payments ($S_t q_t$ in total) are made in the next period. Changing the bond type to zero-coupon bonds with selling price $\frac{1}{\Gamma} q_t$ will not change the essence of the result.

53When the friction (B.8) does not apply, financiers can hold unlimited long or short positions of government bonds, arbitraging away any spread between $S_t$ and $S_t^*$. 

C General model

This section lays out the general model that we use for the calibration and quantitative exercises in Sections 3 and 4. To make exposition easier and to highlight economic mechanisms, a special case of this general model is used in Section 2. In the following sections of the Appendix, we show derivations and proofs for the general model, which also apply to the special case.

This model is more general than the model presented in Section 2 in two ways:

1. the constrained regions face a debt limit $b_{i,t+1}^o \leq B^o$; and for $\beta$ and $B^o \geq 0$ small enough (Assumption E.1 stated in Appendix E), the debt limit constraint is always binding; $b_{i,t+1}^o = B^o$.

2. the transfer function contains a negotiation parameter $\epsilon$, such that region $i$ receives total transfer

$$T_{i,t} = \tau_i b_{i,t} + (1-\epsilon) \tau_i (\bar{b}_t - b_{i,t}) + T_t^u$$

where $b_{i,t}$ is region $i$’s debt stock and $\bar{b}_t = \mu b_t^o + (1-\mu) B^o$ is the cross-region average debt.

The model in Section 2 is a special case when $B^o = 0$ and $\epsilon = 1$. Below we use highlights to indicate the differences from this model.

**Consolidated government’s problem.** Given $(b_0^n, d_0), b_0^o = B^o$ and the interest rate schedules, the consolidated government’s problem consists of choosing a sequence of debt and spending $\{b_{i+1}^n, b_{i+1}^o, d_{i+1}, g_{i,t}, g_{i,t}^o, c_t\}_{t=0}^\infty$ that solves

$$\max_{\{b_{i+1}^n, b_{i+1}^o, d_{i+1}, g_{i,t}, g_{i,t}^o, c_t\}_{t=0}^\infty} \sum (1-\theta) \left[ \mu u(g_{i}^n) + (1-\mu) u(g_{i}^o) \right] + \theta v(c_t)$$

subject to the consolidated budget constraint and constrained regions’ debt limit

$$c_t + \mu g_{i}^n + (1-\mu) g_{i}^o + d_t + \mu b_{i}^n + (1-\mu) b_{i}^o = e + \frac{d_{i+1}}{S(d_{i+1})} + \beta b_{i+1}^n + (1-\mu) \frac{b_{i+1}^o}{R(b_{i+1}^o)}$$

$$b_{i+1}^o \leq B^o$$

**Without a consolidated government:**

**Constrained local government** is subject to the debt limit $b_{i,t+1}^o \leq B^o$, which binds given Assumption E.1. So the constrained region’s debt $b_{i,t+1}^o = B^o$, and we do not need to keep track of it as a state variable. The constrained region’s spending is

$$g_{i,t}^o = f + T_{i,t}^u + \frac{B^o}{R(B^o)} + \tau_i B^o + \tau_i (1-\epsilon) (\bar{b}_t - B^o) - B^o \quad \text{(C.1)}$$

**Unconstrained local government’s problem** is similar to the model in Section 2 except for the modified transfer function. Given the aggregate states $(b_{i}^o, d_{i})$ and individual state $b_{i,t}^o$, an unconstrained local government $i$ chooses its spending $g_{i,t}^o$ and debt $b_{i,t+1}^o$ to maximize the welfare of its residents, taking the sequences of central government policies and interest rates as given. Its problem written recursively is

$$W(b_{i,t}^n; b_{i,t}^o, d_{i}) = \max_{b_{i,t+1}^n, g_{i,t}^o} (1-\theta) u(g_{i,t}^n) + \theta v(c_t) + \beta W(b_{i,t+1}^n; b_{i,t+1}^o, d_{i})$$

subject to the budget constraint

$$g_{i,t}^o + b_{i,t}^o \leq f + \frac{b_{i,t+1}^n}{R(b_{i,t+1}^n)} + \tau_i b_{i,t}^o + \tau_i (1-\epsilon) (\bar{b}_t - b_{i,t}^o) + T_{i,t}^u \quad \text{(C.2)}$$

where $\bar{b}_t = \mu b_t^o + (1-\mu) B^o$. Its optimal choice of $b_{i,t+1}^o$ is characterized by the Euler equation,

$$u_g(g_{i,t}^o) = \beta R(b_{i,t+1}^o) (1-\epsilon) u_g(g_{i,t+1}) \quad \text{(C.3)}$$
Central government’s problem (in a Markov or Ramsey equilibrium) is similar to the model in Section 2, but with the modified local governments’ budget constraints (C.2) and (C.1) and the central government’s budget constraint:

\[ c_t \leq e - f + \frac{d_{t+1}}{S(d_{t+1})} - d_t - \tau_t \tilde{b}_t - T_t^u \quad \text{(C.4)} \]

Using the modified budget constraints to define the spending of the representative local and central governments as functions of transfer and debt, as in Section 2.4:

Unconstrained local: \[ G^n(b^n, b'^n, \tau, T^u) = f + \frac{b'^n}{R(b'^n)} - (1 - \epsilon \tau)b^n + \tau(1 - \epsilon)\tilde{b} + T^u \quad \text{(C.5)} \]

Constrained local: \[ G^o(\tau, T^u) = f + T^u + \frac{B^o}{R(B^o)} - (1 - \epsilon \tau)B^o + \tau(1 - \epsilon)\tilde{b} \quad \text{(C.6)} \]

Central: \[ C(b^n, d, d', \tau, T^u) = e - f + \frac{d'}{S(d')} - d - \tau(\mu b^n + (1 - \mu) B^o) - T^u \quad \text{(C.7)} \]

A Markov-perfect equilibrium in the general model consists of a value function \( V \), central government’s policy rules \( \{ \Phi^c, \Phi^T, \Phi^d \} \), and a policy function \( \Phi^b \) for the unconstrained local government’s debt, such that for all aggregate states \( (b^n, d) \), \( \tau = \Phi^c(b^n, d), T^u = \Phi^T(b^n, d), d' = \Phi^d(b^n, d) \) and \( b'^n = \Phi^b(b^n, d) \) solve

\[
\max_{\tau, T^u, b'^n, d'} \left( 1 - \theta \right) \left[ \mu u \left( G^n(b^n, b'^n, \tau, T^u) \right) + (1 - \mu) u \left( G^o(\tau, T^u) \right) \right] + \theta v \left( C(b^n, d, d', \tau, T^u) \right) + \beta V(b'^n, d')
\]

subject to the representative unconstrained region’s Euler equation and a policy constraint,

\[
\frac{u_g \left( G^n(b^n, b'^n, \tau, T^u) \right)}{T^u} = \beta R(b'^n) \left( 1 - \epsilon \Phi^c(b'^n, d') \right) u_g \left( G^n(b'^n, \Phi^b(b'^n, d'), \Phi^c(b'^n, d'), \Phi^d(b'^n, d'), \Phi^T(b'^n, d')) \right) \text{C.8}
\]

and the central government’s value function satisfies the functional equation

\[
V(b^n, d) = (1 - \theta) \left[ \mu u \left( G^n(b^n, \Phi^b(b^n, d), \Phi^c(b^n, d), \Phi^T(b^n, d)) \right) \right] + (1 - \mu) u \left( G^o(\Phi^c(b^n, d), \Phi^T(b^n, d)) \right) \]

\[ + \theta v \left( C(b^n, d, \Phi^d(b^n, d), \Phi^c(b^n, d), \Phi^T(b^n, d)) \right) + \beta V(\Phi^b(b^n, d), \Phi^d(b^n, d)) \]

The Ramsey problem: Given initial debt positions \( (b^n_0, d_0) \), the optimal (Ramsey) policy of a central government with commitment consists of a sequence of debt and transfer policies \( \{ \tau_t, T^u_t, b^n_{t+1}, d_{t+1} \}_{t=0}^{\infty} \) that solves

\[
\max \sum_{t=0}^{\infty} \beta^t \left\{ (1 - \theta) \left[ \mu u \left( G^n(b^n_t, b^n_{t+1}, \tau_t, T^u_t) \right) + (1 - \mu) u \left( G^o(\tau_t, T^u_t) \right) \right] + \theta v \left( C(b^n_t, d_t, d_{t+1}, \tau_t, T^u_t) \right) \right\}
\]

subject to the following constraints for all time \( t \geq 0 \)

\[
\frac{u_g \left( G^n(b^n_t, b^n_{t+1}, \tau_t, T^u_t) \right)}{T^u_t} = \beta R(b^n_{t+1}) \left( 1 - \epsilon \tau_{t+1} \right) u_g \left( G^n(b^n_{t+1}, b^n_{t+2}, \tau_{t+1}, T^u_{t+1}) \right)
\]

where the budget constraints of central and local governments are implicit in the spending functions \( G^n, G^o \) and \( C \).
D Derivations

We present all derivations for the general model with debt limit \( b^c_{i,t} \leq B^o \), \( B^o \geq 0 \) faced by the constrained regions, and \( \epsilon \in (0, 1] \) parameter in the transfer function. The model presented in Section 2 is a special case when \( B^o = 0 \) and \( \epsilon = 1 \).

D.1 Markov optimality conditions and GEEs

This section derives the Markov optimality conditions (14)–(17) and Generalized Euler Equations (GEEs) for the general case \( B^o \geq 0 \) and \( \epsilon \in (0, 1] \). For expositional convenience, denote the unconstrained regions’ Euler equation (12) by \( \eta(b^n, b^{n'}, d', \tau, T^n) = 0 \), and use \( \eta_{b^n}, \eta_{b^{n'}}, \eta_d, \eta_T \) and \( \eta_{T'} \) to denote partial derivatives.

\[
\begin{align*}
\{ \tau \} & \quad (1 - \theta)[\mu u^n_G G^n + (1 - \mu) u^n_G C^n_T] + \theta v_c C_T = \lambda \eta_T \quad \text{(D.1)} \\
\{ T^n \} & \quad (1 - \theta)[\mu u^n_G G^n_T + (1 - \mu) u^n_G C^n_G^n_T] + \theta v_c C_T = \lambda \eta_T - \zeta \quad \text{(D.2)} \\
\{ b^n \} & \quad (1 - \theta)\mu u^n_{G^n} G^n_{b^n} + \beta v' = \lambda \eta_{b^n} \quad \text{(D.3)} \\
\{ d' \} & \quad \theta v_c C_d + \beta v' = \lambda \eta_d \quad \text{(D.4)}
\end{align*}
\]

where \( \zeta \geq 0 \) is the Lagrange multiplier on the constraint (13). Envelope conditions are

\[
\begin{align*}
V_{b^n} & = (1 - \theta)\mu u^n_{G^n} G^n_{b^n} + (1 - \theta)(1 - \mu) u^n_{G^n} C^n_T + \theta v_c C_T + \lambda \eta_T \Phi_{b^n} + \lambda \eta_{b^n} \Phi_{b^n} + \lambda \eta_d \Phi_{b^n} + \lambda \eta_{T'} \Phi_{b^n} + (\lambda \eta_T - \zeta) \Phi_{b^n} \\
V_{d} & = (1 - \theta)(1 - \mu) u^n_{G^n} G^n_T + \theta v_c C_T + \lambda \eta_T \Phi_{d} + \lambda \eta_{b^n} \Phi_{d} + \lambda \eta_d \Phi_{d} + (\lambda \eta_T - \zeta) \Phi_{d}
\end{align*}
\]

Differentiate \( \eta \) with respect to \( b^n \) and \( d' \),

\[
\begin{align*}
\eta_{b^n} + \eta_T \Phi_{b^n} + \eta_{b^n} \Phi_{b^n} + \eta_d \Phi_{b^n} + \eta_T \Phi_{b^n} &= 0 \\
\eta_T \Phi_d + \eta_{b^n} \Phi_d + \eta_d \Phi_d + \eta_T \Phi_d &= 0
\end{align*}
\]

substitute into \( V_{b^n} \) and \( V_d \),

\[
\begin{align*}
V_{b^n} &= (1 - \theta)\mu u^n_{G^n} G^n_{b^n} + (1 - \theta)(1 - \mu) u^n_{G^n} G^n_T + \theta v_c C_T - \lambda \eta_{b^n} - \zeta \Phi_{b^n} \\
V_d &= \theta v_c C_d - \zeta \Phi_d
\end{align*}
\]

substitute into the last two FOCs,

\[
\begin{align*}
(1 - \theta)\mu u^n_{G^n} G^n_{b^n} + \beta \left[ (1 - \theta)[\mu u^n_{G^n} G^n_{b^n} + (1 - \mu) u^n_{G^n} C^n_{b^n}] + \theta v_c C_{b^n} \right] &= \lambda \eta_{b^n} + \beta(\lambda \eta_{b^n} + \zeta \Phi_{b^n}) \quad \text{(D.5)} \\
\theta v_c C_d + \beta \theta v_c C_d &= \lambda \eta_d + \beta \zeta \Phi_d \quad \text{(D.6)}
\end{align*}
\]

Expanding the auxiliary functions in (D.1), (D.2), (D.5) and (D.6), we get the optimality conditions of
the Markov equilibrium:

\[ \begin{align*}
\{\tau\} & \quad (1 - \theta) \left( \mu u^n g [eb^n + (1 - \epsilon)b] + (1 - \mu) u^n g [eB^n + (1 - \epsilon)b] \right) - \theta v_c [\mu b^n + (1 - \mu) B^n] \\
& = \lambda u^n g [eb^n + (1 - \epsilon)b] \\
\{T^n\} & \quad (1 - \theta) [\mu u^n g + (1 - \mu) u^n g] - \theta v_c = \lambda u^n g - \zeta \\
\{b^n\} & \quad (1 - \theta) \mu u^n g / R(b^n) - \beta \left[ (1 - \theta) \mu u^n g (1 - \epsilon \tau') + \theta v'_c \mu \tau' \right] \\
& = \lambda \left[ u^n g / R(b^n) + \beta R(b^n) (1 - \epsilon \tau') u^n g (1 - \epsilon \tau') \right] + \lambda \Omega_{\delta,\theta} - \beta \lambda u^n g (1 - \epsilon \tau') + \beta \zeta \Phi_{\delta,\theta} \tau' \\
\{d'\} & \quad \theta v_c / S(d') - \beta \theta v'_c = \lambda \Omega_{\delta,\theta} + \beta \zeta \Phi_{\delta,\theta} \tau' \\
\end{align*} \]

where

\[ \Omega_{\delta,\theta} = \beta \epsilon R(b^n) \Phi_{\delta,\theta} x u^n g - \beta R(b^n) (1 - \epsilon \tau') u^n g \left[ (eb^n + (1 - \epsilon)b) \Phi_{\delta,\theta} + \Phi_{\delta,\theta} + \Phi_{\delta,\theta} / R(b^n) \right], \quad x = \{b^n, d\} \]

Setting \( B^n = 0 \) and \( \epsilon = 1 \) reduces the conditions to (14)–(17) of Section 2.4.

Next, we combine the optimality conditions to get the GEEs. From Proposition 2, the constraint on \( T^n \) is binding, so \( \Phi_{\delta,\theta} = \Phi_{\delta,\theta} = 0 \) and the \( \zeta \) terms drop out in (D.5) and (D.6). From (D.1),

\[ \lambda = \frac{1}{\eta} \left( (1 - \theta) [\mu u^n g G^n + (1 - \mu) u^n g G^n] + \theta v_c C_\tau \right) \]

Substitute the expression for \( \lambda \) into (D.5) and (D.6), we get the two GEEs,

\[
\begin{align*}
\mu u^n g G_{\delta,\theta} + \beta \left[ \mu u^n g G_{\delta,\theta} + (1 - \mu) u^n g G_{\delta,\theta} + \frac{\theta}{1 - \theta} v_c C_\delta \right] &= \frac{\eta_{\delta,\theta}}{\eta} \left[ \mu u^n g G_{\tau} + (1 - \mu) u^n g G_{\tau} + \frac{\theta}{1 - \theta} v_c C_\tau \right] + \beta \frac{\eta_{\delta,\theta}}{\eta} \left[ \mu u^n g G_{\tau} + (1 - \mu) u^n g G_{\tau} + \frac{\theta}{1 - \theta} v_c C_\tau \right] \\
\frac{\theta}{1 - \theta} v_c C_\delta + \beta \frac{\theta}{1 - \theta} v_c C_\delta &= \frac{\eta_{\delta,\theta}}{\eta} \left[ \mu u^n g G_{\tau} + (1 - \mu) u^n g G_{\tau} + \frac{\theta}{1 - \theta} v_c C_\tau \right] \\
\end{align*}
\]
The auxiliary functions are

\[
\begin{align*}
G_{b''}^n &= -1 + \epsilon \tau \\
G_{b''}^\tau &= \epsilon b^n + (1 - \epsilon) b \\
G_{T}^n &= 1 \\
G_{b''T}^n &= \frac{1}{R(b^n)} \\
G_{b''}^\tau &= 0 \\
G_{\tau}^n &= \epsilon B^n + (1 - \epsilon) b \\
G_{\tau}^T &= 1 \\
C_{b''} &= -\mu \tau \\
C_{\tau} &= -[\mu b^n + (1 - \mu) B^n] \\
C_{T} &= -1 \\
C_{d} &= -1 \\
C_{d''} &= \frac{1}{S(d''')} \\
\eta_{b''} &= u_{gg} G_{b''}^n \\
\eta_{\tau} &= u_{gg} G_{\tau}^n \\
\eta_{T} &= u_{gg} G_{T}^n \\
\eta_{b'''} &= u_{gg} G_{b'''}^n + \beta \epsilon R(b^n) \Phi_{b'''} u_{g}^n - \beta R(b^n) (1 - \epsilon') u_{g}^n [G_{b'''}^n + G_{\tau}^n \Phi_{b'''} + G_{T}^n \Phi_{b'''} + G_{b'''}^n \Phi_{b'''}] \\
\eta_{d'''} &= \beta \epsilon R(b^n) \Phi_{d'''} u_{g}^n - \beta R(b^n) (1 - \epsilon') u_{g}^n [G_{\tau}^n \Phi_{d'''} + G_{T}^n \Phi_{d'''} + G_{b'''}^n \Phi_{b'''}]
\end{align*}
\]

Note that because individual local government treats the average local debt \( \bar{b} \) as given, there is no additional term in \( G_{b''}^n \) from the effect of \( b^n \) on \( \bar{b} \). Some expressions are slightly different in the extension of Section 4.3.1, because the governments internalize the slopes of their interest rate schedules (captured by highlighted parts):

\[
\begin{align*}
G_{b''}^{n'} &= \frac{1}{R(b^n)} \frac{b^n' R_{b}(b^n)}{R(b^n)^2} \\
C_{d'''} &= \frac{1}{S(d''')} \frac{d'' S_{d}(d''')} {S(d''')^2} \\
\eta_{b''} &= u_{gg} G_{b''}^n \left[ 1 - \frac{b^n' R_{b}(b^n)}{R(b^n)} \right] \\
\eta_{\tau} &= u_{gg} G_{\tau}^n \left[ 1 - \frac{b^n' R_{b}(b^n)}{R(b^n)} \right] \\
\eta_{T} &= u_{gg} \left[ 1 - \frac{b^n' R_{b}(b^n)}{R(b^n)} \right] \\
\eta_{b'''} &= u_{gg} G_{b'''}^n \left[ 1 - \frac{b^n' R_{b}(b^n)}{R(b^n)} \right] - u_{g}^n \left[ \frac{R_{b}'(b^n) + b^n' R_{b}'(b^n)}{R(b^n)^2} - \frac{b^n' R_{b}''(b^n)}{R(b^n)^2} \right] \cdots \\
&= -\beta R_{b}(b^n)' \left( 1 - \epsilon' \right) u_{g}^n + \beta \epsilon R(b^n) \Phi_{b'''} u_{g}^n - \beta R(b^n) (1 - \epsilon') u_{g}^n [G_{b'''}^n + G_{\tau}^n \Phi_{b'''} + G_{T}^n \Phi_{b'''} + G_{b'''}^n \Phi_{b'''}]
\end{align*}
\]
D.1.1 Economic intuition from Markov GEEs

The GEEs derived here provide another way to understand the intuition behind Proposition 3. Rewriting (D.11):

\[
0 = \left( \mu u_g^o b^n + (1 - \mu) u_g^o B^o + (1 - \epsilon) \mu (1 - \mu) (b^n - B^o) (u_g^o - u_g^n) - \frac{\theta}{1 - \theta} v_c [\mu b^n + (1 - \mu) B^o] \right) \tag{D.13}
\]

This condition characterizes the marginal welfare effects of increasing the transfer rate \( \tau \). The signs in parentheses in the GEEs are based on calibrated values of Section 3. Using Implicit Function Theorem, ratios of \( \eta \)-derivatives have the interpretation of marginal effects holding other variables of the local government’s Euler equation constant. For example, \( -\eta_{\tau} / \eta_{b^n'} \) gives the effect of changing \( \tau \) on \( b^n' \) holding constant \( b^n \) and \( d' \).

The right-hand side of (D.13) comprises of three marginal welfare effects of increasing the transfer rate \( \tau \). First, intra-temporally, a higher transfer \( \tau \) shifts spending from central to local governments in the current period. Second, \( \tau \) changes \( b^n' \) (by \( -\eta_{\tau} / \eta_{b^n'} < 0 \) amount), which shifts spending inter-temporally. Third, \( \tau \) changes \( \tau' \) through changing \( b^n' \), which affects within-period spending sharing between central and local governments in the next period.

The consolidated government achieves perfect resource sharing: \( u_g^o = u_g^n = \theta / (1 - \theta) v_c \) (MULG = MUCG). This essentially corresponds to the first line of (D.13). In addition, at the steady state, the third line of (D.13) also collapses into the perfect resource sharing condition. This leaves the second line of (D.13) as the major difference between the consolidated government allocation and the non-commitment (Markov) equilibrium. Specifically, the second line demonstrates that the central government’s marginal value of increasing \( b^n' \) is different from the marginal value perceived by local governments. To see this, we decompose the inter-temporal effect of raising \( b^n' \) (second line of (D.13)) into three parts: the marginal benefit from higher local spending today, the marginal cost of lower local spending tomorrow, and the additional marginal cost to the central

\[\text{Rewriting the other GEE (D.12):}\]

\[
0 = \left( \mu u_g^o b^n + (1 - \mu) u_g^o B^o + (1 - \epsilon) \mu (1 - \mu) (b^n - B^o) (u_g^o - u_g^n) - \frac{\theta}{1 - \theta} v_c [\mu b^n + (1 - \mu) B^o] \right) \tag{D.14}
\]

which captures the marginal welfare effects of increasing \( \tau \), through changing \( d' \) while holding \( b^n' \) constant. For the purpose here we only need to focus on (D.13).
government from larger transfers tomorrow:

\[
\mu u^n - \frac{1}{R(b^n')} - \beta \mu \left( u^n' (1 - \epsilon \tau') + \frac{\theta}{1 - \theta} v_c \tau' \right) \\
\equiv \mu \left( \frac{u^n}{R(b^n')} - \frac{\beta u^n'}{1 - \theta} v_c \tau' \right) \\
MB \text{ today} \\
MC \text{ tomorrow} \\
\text{additional MC: larger transfers tomorrow} \\
\]

(D.15)

The first two terms on the right-hand side of (D.15) constitute the unconstrained local government’s Euler equation and capture the local government’s trade-offs of increasing \( b^n' \). The central government, by contrast, internalizes an additional cost of increasing \( b^n' \): it raises the total transfers paid out tomorrow. When \( \tau' = 0 \), as in the consolidated government case, (D.15) coincides with the unconstrained local government’s Euler equation and is zero, and so at the steady state (D.13) implies perfect resource sharing. When \( \tau' > 0 \), as in the Markov equilibrium, the distortionary transfer drives a wedge between the central and local governments’ perceived marginal values of \( b^n' \): while the local governments think the borrowing cost is cut by \( \tau' > 0 \), the central government recognizes that larger borrowing implies larger transfers tomorrow, and the total borrowing cost paid by the entire fiscal system is still \( R(b^n') \). Thus the central government wants a lower \( b^n' \) than the level determined by the unconstrained local government’s inter-temporal condition.

To achieve this, the central government has to use higher distortionary transfers to relax the unconstrained local government’s budget constraint in the current period and reduce its need to borrow. In the equilibrium, the transfer level is higher than that implied by the within-period term of (D.13) alone. As a result, the weighted marginal utility of local spending is too low relative to that of central spending, \( \mu u^n + (1 - \mu) u^n' < \theta / (1 - \theta) v_c \). In other words, compared with the consolidated government allocation, which is the efficient allocation, the central government over-transfers (Proposition 3).

The over-transfer, however, does not reduce local government debt in equilibrium, because of the ex ante overborrowing effect. When the unconstrained local governments anticipate a larger transfer today, they borrow more in the previous period. Because the central government in this time-consistent equilibrium takes ex ante incentives as forgone, it does not take into account this ex ante effect of larger transfers. The larger transfer, aimed to lower local debt, ends up raising local debt. In other words, over-transfer exacerbates over-borrowing.

19
D.2 Ramsey optimality conditions

This section derives the Ramsey optimality conditions for the general case $B^o \geq 0$ and $\epsilon \in (0,1]$. Denote the unconstrained local government’s Euler equation at time $t$ by $\tilde{\eta}(b^n_t, b^n_{t+1}, \tau_t, T^n_t, b^n_{t+2}, \tau_{t+1}, T^n_{t+1}) = 0$ with the Lagrange multiplier $\beta^t \gamma_t$. Let $\beta^t \zeta^t$ be the Lagrange multiplier on the constraint $T^n_t \geq \tilde{T}$. The time-$t$ optimality conditions for $t > 0$ are:

$$\{\tau_t\} \quad \beta^t(1-\theta)[\mu u^n_t G^n_{g,t} + (1-\mu)u^n_t G^n_{\theta,t}] + \beta^t \theta v_{c,t} C_{r,t} = \beta^t \gamma_t \tilde{\eta}_{r,t} + \beta^t \gamma_{t-1} \tilde{\eta}_{r,t-1}$$

$$\{T_t^n\} \quad \beta^t(1-\theta)[\mu u^n_t G^n_{g,t} + (1-\mu)u^n_t G^n_{\theta,t}] + \beta^t \theta v_{c,t} C_{T,t} = \beta^t \gamma_t \tilde{\eta}_{T,t} + \beta^t \gamma_{t-1} \tilde{\eta}_{T,t-1} - \beta^t \zeta_t$$

$$\{b^n_{t+1}\} \quad \beta^t(1-\theta)\mu u^n_t G^n_{b^n,t} + \beta^t \theta v_{c,t} C_{b^n,t+1} + \beta^{t+1}[(1-\theta)\mu u^n_{b^n+1} G^n_{b^n,t+1} + \theta v_{c,t+1} C_{b^n,t+1}]$$

$$= \beta^t \gamma_t \tilde{\eta}_{b^n,t} + \beta^{t+1} \gamma_{t+1} \tilde{\eta}_{b^n,t+1} + \beta^t \gamma_{t-1} \tilde{\eta}_{b^n,t-1}$$

$$\{d_t+1\} \quad \beta^t \theta v_{c,t} C_{d,t+1} + \beta^{t+1} \theta v_{c,t+1} C_{d,t+1} = 0$$

Expanding the auxiliary functions and writing the above conditions recursively:

$$\{\tau_t\} \quad (1-\theta)(\mu u^n g [eb^n + (1-\epsilon) \tilde{b}] + (1-\mu)u^n g [\epsilon B^o + (1-\epsilon) \tilde{b}]) - \theta v_c [\mu b^n + (1-\mu)B^o]$$

$$= \gamma u^n_{g g} [eb^n + (1-\epsilon) \tilde{b}] + \gamma R(b^n) (\epsilon u^n g - (1-\epsilon) \mu u^n g [eb^n + (1-\epsilon) \tilde{b}])$$

$$\{T_t^n\} \quad (1-\theta)[\mu u^n g + (1-\mu)u^n g] - \theta v_c = \gamma u^n_{g g} - \gamma R(b^n)(1-\epsilon)u^n_{g g} - \zeta^t$$

$$\{b^n\} \quad (1-\theta)\mu u^n_t / R(b^n) - \beta[(1-\theta)\mu u^n_t (1-\epsilon^t) + \theta v_c \mu \epsilon^t]$$

$$= \gamma \left[ u^n_{g g} / R(b^n) + \beta R(b^n) (1-\epsilon^t)u^n_{g g} (1-\epsilon^t) \right]$$

$$- \gamma R(b^n) (1-\epsilon)u^n_{g g} / R(b^n) - \beta \gamma u^n_{g g} (1-\epsilon^t)$$

$$\{d^t\} \quad \theta v_c / S(d^t) - \beta \theta v_c = 0$$

Setting $B^o = 0$ and $\epsilon = 1$, we get conditions (23)–(26) in Section 2.5.

The highlighted parts are the additional terms applicable only when local and central governments internalize the slopes of their interest rate schedules (Section 4.3.1).
\[ \eta_{\nu, t} = u^n_{g, t} G^n_{\nu, t} \left[ 1 - \frac{b^n_{t+1} R_0(b^n_{t+1})}{R(b^n_{t+1})} \right] - v^n_{g, t} \left[ \frac{R_0(b^n_{t+1}) + b^n_{t+1} R_0(b^n_{t+1}) - b^n_{t+1} R_0(b^n_{t+1})}{R(b^n_{t+1})} \right] \cdots \]

\[ \eta_{\nu, t+1} = -\beta R(b^n_{t+1})(1 - \epsilon \tau t) u^n_{g, t+1} - \beta R(b^n_{t+1})(1 - \epsilon \tau t) u^n_{g, t+1} G^n_{\nu, t+1} \]

\[ \eta_{\tau, t} = u^n_{g, t} G^n_{\tau, t} \left[ 1 - \frac{b^n_{t+1} R_0(b^n_{t+1})}{R(b^n_{t+1})} \right] \]

\[ \eta_{\tau, t+1} = \beta R(b^n_{t+1}) u^n_{g, t+1} - \beta R(b^n_{t+1})(1 - \epsilon \tau t) u^n_{g, t+1} G^n_{\tau, t+1} \]

\[ \eta_{T, t} = u^n_{g, t} G^n_{T, t} \left[ 1 - \frac{b^n_{t+1} R_0(b^n_{t+1})}{R(b^n_{t+1})} \right] \]

\[ \eta_{T, t+1} = -\beta R(b^n_{t+1})(1 - \epsilon \tau t) u^n_{g, t+1} G^n_{T, t+1} \]

The model presented in Section 2 is a special case with \( B^0 = 0 \) and \( \epsilon = 1 \) in the above auxiliary functions.
E Proofs

We present all proofs for the general model with debt limit $b_{i,t}^o \leq B_o$, $B_o \geq 0$ faced by the constrained regions and the parameter $\epsilon \in (0,1]$ in the transfer function. **The model presented in Section 2 is a special case when** $B_o = 0$ **and** $\epsilon = 1$.

**Preamble**

The general model requires a modified assumption that replaces Assumption 3(b).

**Assumption E.1.** $\beta$ and $B^o$ are small enough such that given $\epsilon \in (0,1]$, $\beta R(B^o)(1 - \epsilon \bar{\tau}) < 1$, where $\bar{\tau} < 0$ is the lower bound of $\tau$.

This assumption ensures that $b_{i,t}^o \leq B^o$ binds and constrained local governments’ debt is $b_{i,t}^o = B^o$ in equilibrium. In addition, Assumption 4 is restated for the general case as

**Assumption E.2.** (Vertical fiscal imbalance with $B^o$) $f + \bar{T}$ is small relative to $\epsilon$ such that 

$$(1 - \theta)u_g (f + \bar{T} - B^o[1 - 1/R(B^o)]) > \theta v_e (e - f - \bar{T} - \bar{D}(1 - \beta)),$$

where $\bar{D}$ satisfy $S(\bar{D}) = 1/\beta$.

When $B^o = 0$, this modified assumption becomes Assumption 4.

**E.1 Proof of Lemma 1 (Cross-region spending gap in Markov)**

At the steady state, the unconstrained local government’s Euler equation (C.3) implies $\beta R(b_{ss}^o)(1 - \epsilon \tau_{ss}) = 1$. Given the assumptions that $\beta R(B^o)(1 - \epsilon \bar{\tau}) < 1$ (Assumption E.1), $\tau \geq \bar{\tau}$ and $R(\cdot)$ is an increasing function, it is easy to see that

$$b_{ss}^o > B^o,$$

and for the special case $B^o = 0$, this implies $b_{ss}^o > 0$.

It follows then at the steady state,

$$g_{ss}^o = f + T_{ss} + \frac{B^o}{R(B^o)} - (1 - \epsilon \tau_{ss})B^o + \tau_{ss}(1 - \epsilon) \bar{b}_{ss},$$

$$= f + T_{ss} + \left(\frac{1}{R(b_{ss}^o)} - 1 + \epsilon \tau_{ss}\right) B^o + \tau_{ss}(1 - \epsilon) \bar{b}_{ss},$$

$$\geq f + T_{ss} + \left(\frac{1}{R(b_{ss}^o)} - 1 + \epsilon \tau_{ss}\right) B^o + \tau_{ss}(1 - \epsilon) \bar{b}_{ss},$$

$$> f + T_{ss} + \left(\frac{1}{R(b_{ss}^o)} - 1 + \epsilon \tau_{ss}\right) b_{ss}^o + \tau_{ss}(1 - \epsilon) \bar{b}_{ss},$$

$$= g_{ss}^o,$$

where the first inequality follows from $b_{ss}^o > B^o \geq 0$, and the second inequality follows from

$$\frac{1}{R(b_{ss}^o)} - 1 + \epsilon \tau_{ss} = (\beta - 1)(1 - \epsilon \tau_{ss}) < 0,$$

which makes use of the unconstrained local government’s Euler equation (C.3) at the steady state.
E.2 Proof of Proposition 3 (Over-transfer in Markov)

The optimality condition (D.8):

\[(1 - \theta) \left[ \mu u^n_g + (1 - \mu) u^0_g \right] - \theta \nu_c = \lambda u^n_{gg} - \zeta \]

From the Kuhn Tucker Theorem, the multiplier \( \zeta \geq 0 \). Thus the following lemma is sufficient to prove Proposition 3.

**Lemma E.1.** The Lagrange multiplier \( \lambda > 0 \) at the steady state.

*Proof.* The proof will be organized as follows: we first construct an alternate central government’s problem where the implementability constraint (C.8) is relaxed to an inequality (so that the associated Lagrange multipliers are non-negative). Next, we show that the solution to the alternate problem, with non-negative Lagrange multipliers, also solves the original problem around the steady state. Finally, we show that the Lagrange multipliers must be strictly positive.

1. First, we set up the alternate problem:

   **The alternate central government’s problem.** The alternate central government’s problem is similar to the original central government’s problem defined in Section 2.4, except that the implementability constraint is an inequality.

   \[
   \max_{\tau, T^u, b^n', d'} \quad (1 - \theta) \left[ \mu u \left( G^n \left( b^n, b^n', \tau, T^u \right) \right) + (1 - \mu) u \left( G^n \left( \tau, T^u \right) \right) \right] + \theta v \left( C \left( b^n, d, d', \tau, T^u \right) \right) + \beta V \left( b^n', d' \right)
   \]

   subject to,

   \[
   u_g \left( G^n \left( b^n, b^n', \tau, T^u \right) \right) \leq \beta R \left( b^n' \right) \left( 1 - \epsilon \Phi^T \left( b^n', d' \right) \right) u_g \left( G^n \left( b^n', \Phi^b \left( b^n', d' \right), \Phi^\tau \left( b^n', d' \right), \Phi^T \left( b^n', d' \right) \right) \right)
   \]

   where \( T^u \geq \tilde{T} \)

   and the central government’s value function satisfies the functional equation

   \[
   V \left( b^n, d \right) = (1 - \theta) \left[ \mu u \left( G^n \left( b^n, \Phi^b \left( b^n, d \right), \Phi^\tau \left( b^n, d \right), \Phi^T \left( b^n, d \right) \right) \right) + (1 - \mu) u \left( G^n \left( \Phi^b \left( b^n, d \right), \Phi^\tau \left( b^n, d \right), \Phi^T \left( b^n, d \right) \right) \right) \right]
   \]

   \[
   + \theta v \left( C \left( b^n, d, \Phi^b \left( b^n, d \right), \Phi^\tau \left( b^n, d \right), \Phi^T \left( b^n, d \right) \right) \right) + \beta V \left( \Phi^b \left( b^n, d \right), \Phi^d \left( b^n, d \right) \right)
   \]

Denote the Lagrange multiplier of constraint (E.1) by \( \tilde{\lambda} \). The optimality conditions of the alternative problem include conditions (D.7)-(D.10) (after replacing \( \lambda \) with \( \tilde{\lambda} \)) and the complementary slackness condition from the Kuhn-Tucker Theorem,

\[
\tilde{\lambda} \geq 0 \quad \text{and} \quad \tilde{\lambda} \left[ \beta R' \left( 1 - \epsilon \tau' \right) u^n_g - u^n_{gg} \right] = 0
\]

(2) Next, we show that around the steady state, the solution to the alternate problem also satisfies the constraint (E.1) with equality (or equivalently, constraint (C.8) of the general model), so that the solution to the alternate problem also solves the original problem.

We prove it by contradiction. Suppose for contradiction that there is a solution to the alternate problem such that the constraint (E.1) is not binding:

\[
u^n_g < \beta R' \left( 1 - \epsilon \tau' \right) u^n_{gg}
\]

Then from (E.2), we have \( \tilde{\lambda} = 0 \). Plug it into the first order conditions of the alternate problem (still
conditions D.7-D.10), we have

\[(1 - \theta) \left( \mu u^0_g [\epsilon b^n + (1 - \epsilon)b] + (1 - \mu) u^0_g [\epsilon B^n + (1 - \epsilon)b] \right) - \theta v_c [\mu b^n + (1 - \mu) B^n] = 0 \quad (E.4)\]
\[(1 - \theta) \left[ \mu u^0_g + (1 - \mu) u^0_g \right] - \theta v_c = -\zeta \quad (E.5)\]
\[(1 - \theta) \mu u^0_g \frac{1}{R(b^{n\prime})} - \beta (1 - \theta) \mu u^0_g (1 - \epsilon \tau^\prime) = \beta \zeta \Phi^T\theta \quad (E.6)\]
\[\theta v_c / S(d^\prime) - \beta \theta v_c = \beta \zeta \Phi^T\theta \quad (E.7)\]

Define $\hat{\mu} = \frac{\mu [\epsilon b^n + (1 - \epsilon)b]}{\mu b^n + (1 - \mu) B^n}$. We can rewrite (E.4) as

\[(1 - \theta) \left[ \hat{\mu} u^0_g + (1 - \hat{\mu}) u^0_g \right] - \theta v_c = 0 \quad (E.8)\]

From Lemma 1, since $b^n > B^n$ around the steady state, $b^n > \tilde{b} \equiv \mu b^n + (1 - \mu) B^n$ for any $\mu \in (0, 1)$. So we have $\hat{\mu} \in (\mu, 1)$ for any $\epsilon \in (0, 1)$. Further, from Lemma 1, $u^0_g > u^0_g$ around the steady state. Hence the left-hand side of (E.8) is greater than the left-hand side of (E.5), and so $\zeta > 0$, which further implies $T^u = \tilde{T}$.

Because $\tilde{T}$ is not big enough to resolve the vertical fiscal imbalance (Assumption E.2), the distortionary transfer rate $\tau$ must be strictly positive around the steady state, otherwise the total transfers will not be large enough to make (E.5) and (E.8) hold.

$T^u = \tilde{T}$ further implies that $\Phi^T\theta = 0$ and $\Phi_d = 0$. Then we can rewrite (E.6) as

\[(1 - \theta) \mu \left[ u^0_g - \frac{1}{R(b^{n\prime})} - \beta u^0_g (1 - \epsilon \tau^\prime) \right] = \beta \theta \mu v_c \tau^\prime \quad (E.9)\]

With $\tau^\prime > 0$ around the steady state,\(^{55}\) the right-hand side of (E.9) is strictly positive, therefore the left-hand side of (E.9) must also be strictly positive:

\[u^0_g - \beta R(b^{n\prime}) u^0_g (1 - \epsilon \tau^\prime) > 0 \quad (E.10)\]

This contradicts (E.3) when $\tau^\prime > 0$ around the steady state. Hence, the solution of the alternate problem must satisfy (E.1) with equality around the steady state. As such, it also solves the original problem around the steady state if we let $\lambda = \tilde{\lambda}$.

(3) Last, we show that the non-negative $\tilde{\lambda}$ must be strictly positive, hence $\lambda$ is also strictly positive.

Follow the same proof in Step (2), it is easy to see that when $\tilde{\lambda} = 0$, (E.10) has to hold. This contradicts with (E.1). Hence $\tilde{\lambda}$ must be strictly positive.

\[\square\]

### E.3 Proof of Proposition 2 (Minimum uniform transfer in Markov)

Define $\hat{\mu} = \frac{\mu [\epsilon b^n + (1 - \epsilon)b]}{\mu b^n + (1 - \mu) B^n}$. We can rewrite (D.7) as

\[(1 - \theta) \left[ \hat{\mu} u^0_g + (1 - \hat{\mu}) u^0_g \right] - \theta v_c = \frac{\hat{\mu}}{\mu} \lambda u^0_g \quad (E.11)\]

\(^{55}\)Note that both $T^u = \tilde{T}$ and $\tau^\prime > 0$ here follow from the contradiction and are not part of Proposition 3. The properties that $T^u = T$ and $\tau > 0$ around the steady state are formally proved in Proposition 2.
At the steady state,
\[
\lambda u^n_{g g} - \zeta = (1 - \theta) \left[ \mu u^n_{g} + (1 - \mu)u^n_{c} \right] - \theta v_c \\
< (1 - \theta) \left[ \hat{\mu} u^n_{g} + (1 - \hat{\mu})u^n_{c} \right] - \theta v_c \\
= \frac{\hat{\mu}}{\mu} \lambda u^n_{g g} \\
< \lambda u^n_{g g}
\]

The first equality is from (D.8). The first inequality follows because: at the steady state for any \( \mu \in (0, 1) \) and given \( \epsilon \in (0, 1], \hat{\mu} > \mu \) as \( b^n_{ss} > B^o \geq 0 \), and \( u^n_{g} > u^n_{c} \) as \( g^n_{ss} < g^n_{ss} \) (Lemma 1). The second equality follows from (E.11). The second inequality comes from \( \lambda > 0 \) (Proposition 3), \( \hat{\mu} > \mu \) and \( u^n_{g g} < 0 \). Hence, \( \zeta > 0 \), and the constraint \( T^u \geq \bar{T} \) must be binding.

To show \( \tau > 0 \) at the steady state, note that \( u^n_{g} \leq u^n_{c} \) (Lemma 1) and (E.11) together imply
\[
(1 - \theta) u^n_{g} < (1 - \theta) \left[ \hat{\mu} u^n_{g} + (1 - \hat{\mu})u^n_{c} \right] \leq \theta v_c
\]
as \( \lambda > 0 \) and \( u^n_{g g} < 0 \). Since we have established \( T^u = \bar{T} \), if \( \tau \leq 0 \), Assumption E.2 would imply \( (1 - \theta) u^n_{g} > \theta v_c \), which is a contradiction. Therefore, it must be the case that \( \tau > 0 \).

### E.4 Proof of Proposition 4 (Under-transfer in Ramsey)

- In the first part we prove \( \tau = 0 \) at the steady state.

At the steady state, the left-hand side of (D.18) can be simplified to
\[
(1 - \theta) \mu u^n_{g} \frac{1}{R(b^n)} - \beta [(1 - \theta) \mu u^n_{g}(1 - \epsilon \tau) + \theta v_c \mu \tau] = -\beta v_c \mu \tau
\]
where we use the unconstrained region’s Euler equation at the steady state \( \beta R(b^n)(1 - \epsilon \tau) = 1 \).

Similarly, the right-hand side of (D.18) at the steady state becomes
\[
\gamma \left[ u^n_{g g} \frac{1}{R(b^n)} + \beta R(b^n)(1 - \epsilon \tau)u^n_{g g}(1 - \epsilon \tau) \right] - \gamma R(b^n)(1 - \epsilon \tau)u^n_{g g} \frac{1}{R(b^n)} - \beta \gamma u^n_{g g}(1 - \epsilon \tau)
\equiv \gamma \left[ u^n_{g g} \frac{1}{R(b^n)} + u^n_{g g}(1 - \epsilon \tau) - (1 - \epsilon \tau)u^n_{g g} - \beta u^n_{g g}(1 - \epsilon \tau) \right] = 0
\]
where \( \beta R(b^n)(1 - \epsilon \tau) = 1 \) is repeatedly used. Equating the left- and right-hand sides, we can get \( \tau = 0 \) at the steady state.

- Next, we show \( T^u > \bar{T} \) in the Ramsey steady state equilibrium. We first prove the following lemma:

**Lemma E.2.** The Lagrange multiplier \( \gamma > 0 \) at the steady state.\(^{56}\)

*Proof.* By a similar argument as in Section E.2, we can construct an alternate problem (to the Ramsey problem) with inequalities in the implementability constraints and Lagrange multiplier \( \gamma \geq 0 \). Then we can show the solution to this alternate problem also solves the original problem, and so the Lagrange multiplier of the original problem \( \gamma \geq 0 \).

\(^{56}\)Note that \( \gamma > 0 \) holds in this case (despite the absence of distortionary transfer in equilibrium, i.e., \( \tau = 0 \)) because of the horizontal fiscal imbalance between regions. If there is only one type of region (\( \mu = 0 \) or \( \bar{B} = B^o \)), then \( \tau = 0 \) implies \( \gamma = 0 \) and the Ramsey equilibrium would achieve the (efficient) consolidated government allocation.
It remains to show that $\gamma \neq 0$. Suppose for contradiction $\gamma = 0$, then (D.16)-(D.17) become

\[(1 - \theta)\left(\mu u^n_\theta e b^n + (1 - \epsilon)\bar{b} + (1 - \mu)u^n_\theta (B^o + (1 - \epsilon)\bar{b})\right) - \theta v_c [\mu b^n + (1 - \mu)B^o] = 0 \quad \text{(E.12)} \]

\[(1 - \theta)[\mu u^n_\theta + (1 - \mu)u^n_o] - \theta v_c = -\zeta^r(E.13)\]

Define $\tilde{\mu} = \frac{\mu e b^n + (1 - \epsilon)\bar{b}}{\mu b^n + (1 - \mu)B^o}$, then (E.12) can be rewritten as

\[(1 - \theta)[\tilde{\mu} u^n_\theta + (1 - \tilde{\mu})u^n_o] - \theta v_c = 0 \quad \text{(E.14)}\]

Given $\tau = 0$ at the steady state, the unconstrained region's Euler equation implies $\beta R(b^n_{ss}) = 1$, so $b^n_{ss} = \bar{B} > B^o$ (Assumptions 2 and E.1). It follows then $\tilde{\mu} > \mu$ around the steady state for any $\epsilon \in (0,1]$ and $\mu \in (0,1)$. Because $b^n_{ss} > B^o$ and both types of regions get the same transfer $T^u$, we have $g^n_o < g^n_{ss}$ and $u^n_o > u^n_o$. So the left-hand side of (E.14) is greater than the left-hand side of (E.13). This implies $\zeta^r > 0$, and

\[\zeta^r > 0 \text{ also means } T^u = \Bar{T}, \text{ and since we have shown } \tau = 0, \text{ the transfer that local governments get is only } \Bar{T}. \text{ Hence, under the assumption that } f + \Bar{T} \text{ is small relative to } c (\text{Assumption E.2}), \text{ local government spendings } g^n \text{ and } g^o \text{ are low relative to central government consumption } c. \text{ But this contradicts with (E.15). Thus } \gamma > 0. \]

Combining (D.16)--(D.17), we have

\[-(1 - \mu) \left(1 - \frac{\epsilon B^o + (1 - \epsilon)\bar{b}}{eb^n + (1 - \epsilon)\bar{b}}\right)[(1 - \theta)u^n_o - \theta v_c] = \zeta^r + \frac{\gamma - R(b^n)e u^n_o}{eb^n + (1 - \epsilon)\bar{b}} \quad \text{(E.16)}\]

Because $\gamma > 0$ and $\zeta^r \geq 0$, the right hand side of (E.16) is strictly positive.

On the left-hand side of (E.16), because $\tau = 0$ in the Ramsey steady state, from the Euler equation of the unconstrained local government, it must be that $b^n = \bar{B}$ in the Ramsey steady state. Because $\Bar{B} > B^o$ from Assumption E.1, we have $b^n > B^o$ in the Ramsey steady state. Therefore, in order to make the left hand side of (E.16) strictly positive, we need

\[(1 - \theta)u^n_o < \theta v_c \quad \text{(E.17)}\]

Suppose for contradiction $T^u = \Bar{T}$. Given $\tau = 0$, the only transfer to any local government is $\Bar{T}$. Then Assumption E.2 ($f + \Bar{T}$ is small relative to $c$) implies $(1 - \theta)u^n_o > \theta v_c$, which contradicts (E.17). Therefore, it must be the case $T^u > \Bar{T}$. This also implies that $\zeta^r = 0$.

- Lastly, given $\tau = 0$, (D.17) at the steady state becomes

\[(1 - \theta)[\mu u^n_\theta + (1 - \mu)u^n_o] - \theta v_c = \gamma u^n_{gg}(1 - R) - \zeta^r \quad \text{(E.18)}\]

Given $1 - R < 0$, $u^n_{gg} < 0$, $\gamma > 0$ and $\zeta^r = 0$, we have

\[(1 - \theta)[\mu u^n_\theta + (1 - \mu)u^n_o] > \theta v_c.\]
F  Model extensions

F.1 Allowing both types of local governments to borrow

In this subsection, we relax the assumption in the benchmark model that only the unconstrained regions’ local governments can borrow. We assume that both types of regions can borrow, but they face different interest schedules. We show that the constrained regions, which face a higher interest schedule, would have lower local government debt levels than the unconstrained regions.\(^{57}\) The central government will always want to use debt-dependent transfers to subsidize the unconstrained regions (which are more indebted) and the main results in our paper carry through. For exposition purposes, we still use the terms “constrained” and “unconstrained”, although now both types of local governments can borrow freely. In particular, we make the following assumption to replace Assumption 2:

Assumption F.1. Local governments in both types of regions can borrow.

(a) The unconstrained and constrained local governments face the interest schedules \(R^n(\cdot)\) and \(R^o(\cdot)\), respectively. Both \(R^n(\cdot)\) and \(R^o(\cdot)\) are increasing and convex functions, and there exist \(\bar{B}^n\) and \(\bar{B}^o\) such that \(R^n(\bar{B}^n) = R^o(\bar{B}^o) = 1/\beta\);

(b) \(\bar{B}^o\) is sufficiently small relative to \(\bar{B}^n\). A sufficient condition when \(\tau_t \in [0, \tau_{max}]\) is that \(\bar{B}^o < (1 - \tau_{max})\bar{B}^n\).

Assumption F.1(a) allows the constrained and unconstrained regions to borrow freely under different interest schedules. Assumption F.1(b) ensures that the “constrained” regions’ interest schedule is high enough, so that (as we will show later) they have lower debt and higher spending in equilibrium than the “unconstrained” regions. The underlying reasons for the difference in interest schedules may be due to differences in credit risks, targeted investor base or tax treatment. We do not distinguish the specific reasons behind it.

The additional technical difficulty is that when both types of regions can borrow, there is an additional state variable, which is the debt stock of the constrained regions (\(b^o\)). The additional state variable not only increases the dimensionality of the problem, but also brings additional derivative terms of the policy functions, such as the derivatives of the the unconstrained regions’ spending and borrowing choices with respect to the constrained regions’ debt. To keep the model tractable, we make the following simplifying assumption that only the unconstrained regions receive the debt-dependent transfers \(\tau b^n\). As we will show later, this assumption means that we do not need to keep track of the constrained regions’ debt stock \(b^o\) as a state variable in solving the Markov equilibrium.\(^{58}\)

Assumption F.2. Constrained local governments do not face the distortionary transfer scheme \(\tau\).

Although allowing only one type of regions to receive debt-dependent transfers is a technical simplification, Assumption F.2 also has its real-world relevance. For example, Rodden (2005b) documents a similar system in Germany, where under their equalization scheme, only some states are eligible for discretionary transfers. The equalization scheme, which is an intergovernmental redistribution scheme, includes two parts: formula-based transfers and discretionary transfers. The formula-based part divides all German states into two categories based on their tax revenues: those with lower-than-average tax revenues (the recipient states), and those with higher tax revenues (the paying states). Only the recipient states are eligible to receive the

---

\(^{57}\)In the benchmark model presented in Section 2, the heterogeneity in the level of local borrowing is achieved by assuming that the constrained regions cannot borrow, either due to self-prudence or financial market frictions.

\(^{58}\)With Assumption F.2, \(b^o\) is no longer a state variable for the central government’s and the unconstrained regions’ problems, although it is still a state variable for the constrained regions’ problem.
discretionary transfers, which also include the bailouts to debt-distressed states. A paying state would not qualify for the discretionary transfers, even if the state is heavily indebted.

With Assumptions F.1 and F.2, the local governments’ Euler equations are

\[ u_g(y^n_t) = \beta R^n(b^n_{t+1})(1-\tau_t)u_g(y^n_t) \]
\[ u_g(y^o_t) = \beta R^o(b^o_{t+1})u_g(y^o_t) \]

The budget constraints are

Unconstrained local: \[ G^n(b^n, b^{\prime n}, \tau, T^n) = f + \frac{b^{\prime n}}{R(b^{\prime n})} - (1-\tau)b^n + T^n \]

Constrained local: \[ G^o(b^o, b^{\prime o}, \tau, T^o) = f + \frac{b^{\prime o}}{R(b^{\prime o})} - b^o + T^o \]

Central: \[ C(b^n, b^o, d, d', \tau, T^u) = e - f + \frac{d'}{S(d')} - d - \tau b^n - T^u \]

**Definition F.1.** A Markov-perfect equilibrium consists of a value function \( V \), central government’s policy rules \( \{\Phi^r, \Phi^t, \Phi^d\} \), and policy functions \( \{\Phi^b, \Phi^o\} \) for two types of local government’s debt, such that for all aggregate states \( (b^n, b^o, d) \), \( \tau = \Phi^r(b^n, b^o, d) \), \( T^u = \Phi^T(b^n, b^o, d) \), \( d' = \Phi^d(b^n, b^o, d) \), \( b^\prime = \Phi^b(b^n, b^o, d) \) and \( b^{\prime o} = \Phi^o(b^n, b^o, d) \) solve

\[ \max_{\tau, T^u, b^{\prime n}, b^{\prime o}, d'} \left( (1-\theta) \left[ \mu u \left( G^n(b^n, b^{\prime n}, \tau, T^n) \right) + (1-\mu)u \left( G^o(b^o, b^{\prime o}, \tau, T^o) \right) \right] + \theta v \left( C(b^n, b^o, d, d', \tau, T^u) \right) \right) \]

subject to the local regions’ Euler equations and a policy constraint,

\[ u_g \left( G^n(b^n, b^{\prime n}, \tau, T^n) \right) = \beta R^n(b^{\prime n})(1-\Phi^r(b^{\prime n}, b^{\prime o}, d')) \]
\[ u_g \left( G^n(b^n, \Phi^b(b^n, b^{\prime o}, d'), \Phi^r(b^{\prime n}, b^{\prime o}, d'), \Phi^T(b^{\prime n}, b^{\prime o}, d')) \right) \]  
(F.1)

\[ u_g \left( G^o(b^o, b^{\prime o}, \tau, T^o) \right) = \beta R^o(b^{\prime o})u_g \left( G^o(b^{\prime o}, \Phi^o(b^{\prime o}, b^{\prime o}, d'), \Phi^r(b^{\prime n}, b^{\prime o}, d'), \Phi^T(b^{\prime n}, b^{\prime o}, d')) \right) \]  
(F.2)

\[ T^u \geq T \]  
(F.3)

and the central government’s value function satisfies the functional equation

\[ V(b^n, b^o, d) = (1-\theta) \left[ \mu u \left( G^n(b^n, \Phi^b(b^n, b^o, d), \Phi^r(b^n, b^o, d), \Phi^T(b^n, b^o, d)) \right) \right] \]
\[ + (1-\mu)u \left( G^o(b^n, \Phi^o(b^n, b^n, d), \Phi^T(b^n, b^n, d)) \right) \]
\[ + \theta v \left( C(b^n, b^o, d, \Phi^d(b^n, b^o, d), \Phi^r(b^n, b^o, d), \Phi^T(b^n, b^o, d)) \right) \]
\[ + \beta V(\Phi^b(b^n, b^n, d), \Phi^o(b^n, b^n, d), \Phi^d(b^n, b^n, d)) \]

**Derivation of Markov optimality conditions**

Denote the local governments’ Euler equations (F.1) and (F.2) by \( \eta^n(b^n, b^{\prime n}, b^{\prime o}, d', \tau, T^u) = 0 \) and
\( \eta^o(b^o, b'^o, d', \tau, T^n) = 0, \) and the corresponding multipliers by \( \lambda^o \) and \( \lambda^o. \)

\[
\begin{align*}
\{\tau\} & \quad (1 - \theta)\mu u^o G^n_{\theta^n} + (1 - \mu) u^o G^n_{\theta^o} + \theta v_c C_{\tau} = \lambda^o \eta^o_{\theta^n} + \lambda^o \eta^o_{\theta^o} \\
\{T^n\} & \quad (1 - \theta)\mu u^o G^n_{\theta^n} + (1 - \mu) u^o G^n_{\theta^o} + \theta v_c C_T = \lambda^o \eta^o_{\theta^n} + \lambda^o \eta^o_{\theta^o} - \zeta \\
\{b^o\} & \quad (1 - \theta)\mu u^o G^n_{b^o} + \beta V^n_{b^o} = \lambda^o \eta^o_{b^o} + \lambda^o \eta^o_{b^o}' \\
\{b'^o\} & \quad (1 - \theta)(1 - \mu) u^o G^n_{b'^o} + \beta V^n_{b'^o} = \lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o}' \\
\{d'\} & \quad \theta v_c C_{d'} + \beta V^n_{d'} = \lambda^o \eta^o_{d'} + \lambda^o \eta^o_{d'}' 
\end{align*}
\]

where \( \zeta \geq 0 \) is the Lagrange multiplier on the \( T \) constraint (F.3). Envelope conditions are

\[
\begin{align*}
V^n_{b^n} & = (1 - \theta)\left\{ \mu u^o G^n_{b^n} + C^n_{b^n} + G^n_{\theta^n} \Phi^n_{b^n} + G^n_{\theta^o} \Phi^n_{b^o} + G^n_{\theta^o} \Phi^n_{b'^o} + G^n_{\theta^o} \Phi^n_{b'^o} \right\} \\
& + \theta v_c \left[ C^n_{b^n} + C^n_{\tau} \Phi^n_{b^n} + C^n_{d'} \Phi^n_{b^o} + C^n_{d'} \Phi^n_{b'^o} \right] + \beta \left[ V^n_{b^n} \Phi^n_{f^n} + V^n_{b^n} \Phi^n_{f^o} + V^n_{b^n} \Phi^n_{f'^o} \right] \\
V^n_{b'^o} & = (1 - \theta)\left\{ \mu u^o G^n_{b'^o} + C^n_{b'^o} + G^n_{\theta^n} \Phi^n_{b'^o} + G^n_{\theta^o} \Phi^n_{b'^o} + G^n_{\theta^o} \Phi^n_{b'^o} + G^n_{\theta^o} \Phi^n_{b'^o} \right\} \\
& + \theta v_c \left[ C^n_{b'^o} + C^n_{\tau} \Phi^n_{b'^o} + C^n_{d'} \Phi^n_{b'^o} + C^n_{d'} \Phi^n_{b'^o} \right] + \beta \left[ V^n_{b'^o} \Phi^n_{f^n} + V^n_{b'^o} \Phi^n_{f^o} + V^n_{b'^o} \Phi^n_{f'^o} \right] \\
V^n_{d} & = (1 - \theta)\left\{ \mu u^o G^n_{d^n} + C^n_{d^n} + G^n_{\theta^n} \Phi^n_{d^n} + G^n_{\theta^o} \Phi^n_{d^o} + G^n_{\theta^o} \Phi^n_{d'^o} + G^n_{\theta^o} \Phi^n_{d'^o} \right\} \\
& + \theta v_c \left[ C^n_{d^n} + C^n_{\tau} \Phi^n_{d^n} + C^n_{d'} \Phi^n_{d^o} + C^n_{d'} \Phi^n_{d'^o} \right] + \beta \left[ V^n_{d^n} \Phi^n_{f^n} + V^n_{d^o} \Phi^n_{f^o} + V^n_{d'^o} \Phi^n_{f'^o} \right]
\end{align*}
\]

substitute the FOCs into the Envelope conditions,

\[
\begin{align*}
V^n_{b^n} & = (1 - \theta)\left\{ \mu u^o G^n_{b^n} + \theta v_c C^n_{b^n} + (\lambda^o \eta^o_{b^n} + \lambda^o \eta^o_{b^o}) \Phi^n_{b^n} + (\lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o}') \Phi^n_{b^o} + \lambda^o \eta^o_{b'^o} \right\} \\
& + \left( \lambda^o \eta^o_{b^n} + \lambda^o \eta^o_{b^o} \right) \Phi^n_{b^n} + \left( \lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o} \right) \Phi^n_{b^o} + \left( \lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o} \right) \Phi^n_{b'^o} \\
V^n_{b'^o} & = (1 - \theta)(1 - \mu) u^o G^n_{b'^o} + \theta v_c C^n_{b'^o} + (\lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o}) \Phi^n_{b'^o} + (\lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o}') \Phi^n_{b'^o} + \lambda^o \eta^o_{b'^o} \right\} \\
& + \left( \lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o} \right) \Phi^n_{b'^o} + \left( \lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o} \right) \Phi^n_{b'^o} + \left( \lambda^o \eta^o_{b'^o} + \lambda^o \eta^o_{b'^o} \right) \Phi^n_{b'^o} \\
V^n_{d} & = \theta v_c C^n_{d} + (\lambda^o \eta^o_{d^n} + \lambda^o \eta^o_{d^o}) \Phi^n_{d^n} + (\lambda^o \eta^o_{d'^o} + \lambda^o \eta^o_{d'^o}') \Phi^n_{d^o} + (\lambda^o \eta^o_{d'^o} + \lambda^o \eta^o_{d'^o} \right\} \\
& + \left( \lambda^o \eta^o_{d^n} + \lambda^o \eta^o_{d^o} \right) \Phi^n_{d^n} + \left( \lambda^o \eta^o_{d'^o} + \lambda^o \eta^o_{d'^o} \right) \Phi^n_{d^o} + \left( \lambda^o \eta^o_{d'^o} + \lambda^o \eta^o_{d'^o} \right) \Phi^n_{d'^o} 
\end{align*}
\]

Differentiate \( \eta^o \) and \( \eta^o \) with respect to \( b^o, b'^o \) and \( d \),

\[
\begin{align*}
\eta^n_{b^n} + \eta^n_{b^o} \Phi^n_{b^n} + \eta^n_{b'^o} \Phi^n_{b^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} = 0 \\
\eta^n_{b^n} + \eta^n_{b^o} \Phi^n_{b^n} + \eta^n_{b'^o} \Phi^n_{b^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} = 0 \\
\eta^n_{b^o} + \eta^n_{b'^o} \Phi^n_{b^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} = 0 \\
\eta^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} = 0 \\
\eta^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} = 0 \\
\eta^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} + \eta^n_{b'^o} \Phi^n_{b'^o} = 0
\end{align*}
\]

substitute into \( V^n_{b^n}, V^n_{b'^o} \) and \( V^n_{d} \),

\[
\begin{align*}
V^n_{b^n} & = (1 - \theta)\left\{ \mu u^o G^n_{b^n} + \theta v_c C^n_{b^n} + \lambda^o \eta^o_{b^n} \right\} \\
V^n_{b'^o} & = (1 - \theta)(1 - \mu) u^o G^n_{b'^o} + \theta v_c C^n_{b'^o} + \lambda^o \eta^o_{b'^o} \right\} \\
V^n_{d} & = \theta v_c C^n_{d} - \zeta \Phi^n_{d^n}
\end{align*}
\]

29
substitute into the last three FOCs,

$$(1 - \theta)\mu u_\theta^n C_{\theta\theta}^{n\prime} + \beta \left[ (1 - \theta)\mu u_\theta^{n\prime} C_{\theta\theta}^{n\prime\prime} + \theta v_c C_{\theta d}^{n\prime} \right] = \lambda^\theta \eta_{\theta\theta}^{n\prime} + \lambda^\theta \eta_{\theta\theta}^{n\prime\prime} + \beta \lambda^\theta \eta_{\theta d}^{n\prime} + \beta \zeta^\theta \Phi_{\theta d}^{T\prime} \tag{F.9}$$

$$(1 - \theta)(1 - \mu)u_\theta^o C_{\theta\theta}^{o\prime} + \beta \left[ (1 - \theta)(1 - \mu) u_\theta^{o\prime} C_{\theta\theta}^{o\prime\prime} + \theta v_c C_{\theta d}^{o\prime} \right] = \lambda^\theta \eta_{\theta\theta}^{o\prime} + \lambda^\theta \eta_{\theta\theta}^{o\prime\prime} + \beta \lambda^\theta \eta_{\theta d}^{o\prime} + \beta \zeta^\theta \Phi_{\theta d}^{T\prime} \tag{F.10}$$

$$\theta v_c C_{\theta d} + \beta \theta v_c C_{\theta d}^{n\prime} = \lambda^\theta \eta_{\theta d}^{n\prime} + \lambda^\theta \eta_{\theta d}^{o\prime} + \beta \zeta^\theta \Phi_{\theta d}^{T\prime} \tag{F.11}$$

Expanding the auxiliary functions in (F.4)-(F.5) and (F.9)-(F.11), we get the following Markov optimality conditions.

$$\{\tau\} \quad (1 - \theta)[\mu u_\theta^n b^n] - \theta v_c[\mu b^n] = \lambda^n u_{\theta g}^n b^n \tag{F.12}$$

$$\{T^n\} \quad (1 - \theta)[\mu u_\theta^n + (1 - \mu) u_\theta^o] - \theta v_c = \lambda^n u_{\theta g}^n + \lambda^o u_{\theta g}^o - \zeta \tag{F.13}$$

$$\{b^n\} \quad (1 - \theta)\mu u_\theta^n / R^n(b^n') - \beta \left[ (1 - \theta)\mu u_\theta^{n\prime} (1 - \tau') + \theta v_c \mu \tau' \right] = \lambda^n \left[ u_{\theta g}^n / R^n(b^n') + \beta R^n(b^n') (1 - \tau') u_{\theta g}^{n\prime} (1 - \tau') \right] + \lambda^n \Omega^n_{\theta\theta'} + \lambda^o \Omega^n_{\theta\theta'} \tag{F.14}$$

$$\{b^o\} \quad (1 - \theta)(1 - \mu)u_\theta^o / R^o(b^o') - \beta \left[ (1 - \theta)(1 - \mu) u_\theta^{o\prime} \right] = \lambda^o \left[ u_{\theta g}^o / R^o(b^o')} + \beta R^o(b^o') u_{\theta g}^{o\prime} \right] + \lambda^n \Omega^n_{\theta\theta'} + \lambda^o \Omega^n_{\theta\theta'} - \beta \lambda^o u_{\theta g}^{o\prime} + \beta \zeta^o \Phi_{\theta d}^{T\prime} \tag{F.15}$$

$$\{d\} \quad \theta v_c / S(d) - \beta \theta v_c = \lambda^n \Omega^n_{\theta d} + \lambda^o \Omega^o_{\theta d} + \beta \zeta^o \Phi_{\theta d}^{T\prime} \tag{F.16}$$

where

$$\Omega^n_{\theta\theta'} = \beta R^n(b^n') \Phi_{\theta}^{n\prime} u_{\theta}^{n\prime} - \beta R^n(b^n') (1 - \tau') u_{\theta}^{n\prime} \left[ b^n' \Phi_{\theta}^{n\prime} + \Phi_{\theta}^{T\prime} + \Phi_{\theta}^{n\prime} / R^n(b^n') \right], \quad x = \{b^n, d\}$$

$$\Omega^o_{\theta\theta'} = -\beta R^o(b^o') u_{\theta g}^{o\prime} \left[ \Phi_{\theta}^{T\prime} + \Phi_{\theta}^{o\prime} / R^o(b^o') \right], \quad x = \{b^o, d\}$$
The auxiliary functions are

\[
\begin{align*}
G_{b_n}^n &= -(1 - \tau) \\
G_{\tau n}^n &= b^n \\
G_T^n &= 1 \\
G_{b_n}^{o} &= \frac{1}{R^n(b^n)} \\
G_{\tau n}^{o} &= -1 \\
G_T^{o} &= 0 \\
G_{b_n}^{o} &= 1 \\
G_{b_n}^{o} &= \frac{1}{R^n(b^n)} \\
C_{b_n} &= -\mu \tau \\
C_{b_n} &= 0 \\
C_{\tau} &= -\mu b^n \\
C_T &= -1 \\
C_d &= -1 \\
C_{d_f} &= \frac{1}{S(d_f)} \\
\eta_{b_n}^n &= u_{gg} G_{b_n}^n \\
\eta_{\tau n}^n &= \frac{n}{n} G_{\tau n}^n \\
\eta_T^o &= \frac{n}{n} G_T^o \\
\eta_{b_n}^{o} &= u_{gg} G_{b_n}^{o} + \beta R^n(b^n) \Phi_{b_n} \eta_{b_n} + \beta R^n(b^n) (1 - \tau) \eta_{g_{b_n}}^n [G_{b_n}^{o} + G_{\tau n}^{o} \Phi_{b_n} + G_T^{o} \Phi_T + G_{b_n} + \Phi_{b_n}] \\
\eta_{\tau n}^{o} &= \beta R^n(b^n) \Phi_{\tau n} \eta_{\tau n} + \beta R^n(b^n) (1 - \tau) \eta_{g_{\tau n}}^o [G_{\tau n}^{o} \Phi_{\tau n} + G_T^{o} \Phi_T + G_{\tau n} + \Phi_{\tau n}] \\
\eta_T^{o} &= \beta R^n(b^n) \Phi_T \eta_T + \beta R^n(b^n) (1 - \tau) \eta_{g_T}^o [G_T^{o} \Phi_T + G_{b_n} + \Phi_{b_n}] \\
\eta_{b_n}^{o} &= \frac{n}{n} G_{b_n}^{o} = -u_{gg} \\
\eta_{\tau n}^{o} &= \frac{n}{n} G_{\tau n}^{o} = 0 \\
\eta_T^{o} &= \frac{n}{n} G_T^o \\
\eta_{b_n}^{o} &= \beta R^n(b^n) u_{gg} \left[ G_{T}^{o} \Phi_{T}^{o} + G_{b_n}^{o} \Phi_{b_n}^{o} \right] \\
\eta_{\tau n}^{o} &= \beta R^n(b^n) u_{gg} \left[ G_{\tau n}^{o} \Phi_{\tau n}^{o} - 1 + G_T^{o} \Phi_T^{o} + G_{b_n}^{o} \Phi_{b_n}^{o} \right] \\
\eta_T^{o} &= \beta R^n(b^n) u_{gg} \left[ G_{T}^{o} \Phi_T^{o} + G_{b_n}^{o} \Phi_{b_n}^{o} \right]
\end{align*}
\]

Outline of the Proofs of Propositions 2 and 3 when both types of regions can borrow

We first show Lemma 1 \((g_{ss}^b > g_{ss}^b)\).

At the steady state, the local governments' Euler equations \((F.1)\) and \((F.2)\) are

\[
\begin{align*}
\beta R^n(b_n^o)(1 - \tau_{ss}) &= 1 \\
\beta R^n(b_n^o) &= 1
\end{align*}
\]
which implies \( R^n(b^n_{ss})(1 - \tau_{ss}) = R^o(b^o_{ss}) = \frac{1}{\beta} \). It follows then at the steady state,

\[
g^o_{ss} = f + \frac{b^o_{ss}}{R^o(b^o_{ss})} - b^o_{ss} + T^u_{ss} \\
= f - (1 - \beta)b^o_{ss} + T^u_{ss} \\
= f - (1 - \beta)\bar{b}^o + T^u_{ss} \\
g^n_{ss} = f + \left( \frac{1}{R^n(b^n_{ss})} - 1 + \tau_{ss} \right)b^n_{ss} + T^u_{ss} \\
= f - (1 - \beta)(1 - \tau_{ss})b^n_{ss} + T^u_{ss}
\]

Therefore, \( g^o_{ss} > g^n_{ss} \) is equivalent to \( \bar{b}^o < (1 - \tau_{ss})b^n_{ss} \). The latter is guaranteed by Assumption F.1(b).

Next we show Propositions 2 and 3.

It is easy to see that if the central government always finds it optimal to choose \( T^u_{l} = \bar{T} \) in equilibrium (which is to be verified later), the constrained regions’ maximization problem is independent of the central government and the unconstrained regions’ choices. That is, if \( T^u_{l} = \bar{T} \) holds, \( (b^o_{l}, g^o_{ss}) \) will be determined by the constrained regions’ budget constraint and Euler equation rather than \( \tau_l \) and \( b^o_{l} \). Meanwhile, the central government’s choice of \( \tau_l \) and the unconstrained regions’ choice of \( b^n_{l} \) and \( g^n_{ss} \) will be independent of \( b^o_{l} \).

Therefore, as long as \( T^u_{l} = \bar{T} \) in equilibrium, the constrained regions’ Euler equation (F.2) can be dropped from the central government’s problem, \( \Phi^o_{lu} = \Phi^o_{lu} = \Phi^o_{lu} = 0 \) and \( \Phi^o_{lu} = \Phi^o_{lu} = 0 \). Thus, the optimality conditions (F.12)-(F.16) can be simplified to

\[
\begin{align*}
\{\tau\} & \quad (1 - \theta)[\mu u^n_{gg} b^n] - \theta \nu_c[\mu b^n] = \lambda^n u^n_{gg} b^n \quad \text{(F.17)} \\
\{T^u\} & \quad (1 - \theta)[\mu u^n_{gg} + (1 - \mu)u^n_{gg}] - \theta \nu_c = \lambda^n u^n_{gg} - \zeta \quad \text{(F.18)} \\
\{b^o\} & \quad (1 - \theta)\mu u^n_{gg} / R^n(b^n) - \beta \left[ (1 - \theta)\mu u^n_{gg} \left( 1 - \tau' \right) + \theta \nu_c \mu \tau' \right] \\
& \quad = \lambda^n \left[ u^n_{gg} / R^n(b^n) + \beta R^n(b^n) \left( 1 - \tau' \right)^2 u^n_{gg} \right] + \lambda^n \Omega^n_{b^n} - \beta \lambda^n' u^n_{gg} (1 - \tau') + \beta \zeta' \Phi^n_{b^n} \quad \text{(F.19)} \\
\{b^n\} & \quad \nu_c / R^n(b^n) - \beta \nu_c = 0 \quad \text{(F.20)} \\
\{d^l\} & \quad \theta \nu_c / S(d^l) - \beta \theta \nu_c = \lambda^n \Omega^n_{d^l} + \beta \zeta' \Phi^n_{d^l} \quad \text{(F.21)}
\end{align*}
\]

The rest of the proofs of Propositions 2 and 3 follow the proofs in the benchmark model. Intuitively, the vertical fiscal imbalance entails central government’s transfers to regions, either through \( \tau \) or \( T^u \). Because the constrained regions’ debt burden is smaller than the unconstrained regions (in absence of any transfers through \( \tau \)), the constrained regions’ public spending will be higher than the unconstrained regions. Hence the central government always prefers using \( \tau \), which can channel more transfers to the unconstrained regions. That is why the central government finds it optimal to set \( T^u = \bar{T} \).
F.2 Adding adjustment cost on $T^u_t$

In this subsection, we relax the assumption that the uniform transfer is bounded below ($T^u_t \geq \bar{T}$). Instead, we assume that the central government incurs a cost when its uniform transfer deviates from $\bar{T}$, and this cost is quadratic, $\frac{1}{2} \phi(T^u_t - \bar{T})^2$. We show that the main results in the paper continue to hold here.

The budget constraints are

Unconstrained local: $G^n(b^n, b^n', \tau, T^u) = f + \frac{b^n'}{R(b^n')} - (1 - \tau)b^n + T^u$

Constrained local: $G^n(\tau, T^u) = f + T^u + \frac{B^n}{R(B^n)} - (1 - \tau)B^n$

Central: $C(b^n, d, d', \tau, T^u) = e - f + \frac{d'}{S(d')} - d - \tau [\mu b^n + (1 - \mu)B^n] - T^u - \frac{1}{2} \phi(T^u - \bar{T})^2$

**Definition F.2.** A Markov-perfect equilibrium consists of a value function $V$, central government’s policy rules $(\Phi^\tau, \Phi^T, \Phi^d)$, and a policy function $\Phi^b$ for the unconstrained local government’s debt, such that for all aggregate states $(b^n, d, \tau = \Phi^\tau(b^n, d), T^u = \Phi^T(b^n, d), d' = \Phi^d(b^n, d) \text{ and } b'^n = \Phi^b(b^n, d)$ solve

$$\max_{\tau, T^u, b'^n, d'} (1 - \theta) \left[ \mu u \left( G^n(b^n, b'^n, \tau, T^u) \right) + (1 - \mu) u \left( G^n(\tau, T^u) \right) \right] + \theta v \left( C(b^n, d, d', \tau, T^u) \right) + \beta V(b'^n, d')$$

subject to the representative unconstrained region’s Euler equation,

$$u_g \left( G^n(b^n, b'^n, \tau, T^u) \right) = \beta R(b'^n)(1 - \Phi^\tau(b'^n, d'))u_g \left( G^n(b'^n, \Phi^b(b'^n, d'), \Phi^\tau(b'^n, d'), \Phi^T(b'^n, d')) \right)$$

and the central government’s value function satisfies the functional equation

$$V(b^n, d) = (1 - \theta) \left[ \mu u \left( G^n(b^n, \Phi^b(b^n, d), \Phi^\tau(b^n, d), \Phi^T(b^n, d)) \right) + (1 - \mu) u \left( G^n(\Phi^T(b^n, d), \Phi^T(b^n, d)) \right) \right] + \theta v \left( C(b^n, d, \Phi^d(b^n, d), \Phi^\tau(b^n, d), \Phi^T(b^n, d)) \right) + \beta V(b^b, d, \Phi^d(b^n, d))$$

**Derivation of Markov optimality conditions**

Denote the unconstrained local governments’ Euler equation (F.22) by $\eta(b^n, b'^n, d', \tau, T^u) = 0$ and the corresponding multiplier by $\lambda$.

$$\begin{align*}
\{ \tau \} & : (1 - \theta) \mu u_g G^n + (1 - \mu) u_g G^n \tau + \theta v \tau C = \lambda \eta \tau \\
\{ T^u \} & : (1 - \theta) \mu u_g G^n + (1 - \mu) u_g G^n \tau + \theta v \tau C_T = \lambda \eta T \\
\{ b'^n \} & : (1 - \theta) \mu u_g G^n + \beta V'_{b'} = \lambda \eta_{b'^n} \\
\{ d' \} & : \theta v \tau C_{d'} + \beta V'_{d'} = \lambda \eta_{d'}
\end{align*}$$

Envelope conditions are

$$\begin{align*}
V_{b^n} &= (1 - \theta) \mu u_g G^n + \theta v \tau C_{b^n} + \lambda \eta_{\tau} \Phi_{b^n} + \lambda \eta_{b^n} \Phi_{b^n} + \lambda \eta_{d^n} \Phi_{b^n} + \lambda \eta_{T} \Phi_{b^n} \\
V_{d} &= (1 - \theta) \mu u_g G^n + \theta v \tau C_{d} + \lambda \eta_{\tau} \Phi_{d} + \lambda \eta_{b^n} \Phi_{b^n} + \lambda \eta_{d^n} \Phi_{b^n} + \lambda \eta_{T} \Phi_{b^n}
\end{align*}$$

substitute the FOCs into the Envelope conditions,

$$\begin{align*}
V_{b^n} &= (1 - \theta) \mu u_g G^n + \theta v \tau C_{b^n} + \lambda \eta_{\tau} \Phi_{b^n} + \lambda \eta_{b^n} \Phi_{b^n} + \lambda \eta_{d^n} \Phi_{b^n} + \lambda \eta_{T} \Phi_{b^n} \\
V_{d} &= \theta v \tau C_{d} + \lambda \eta_{\tau} \Phi_{d} + \lambda \eta_{b^n} \Phi_{b^n} + \lambda \eta_{d^n} \Phi_{b^n} + \lambda \eta_{T} \Phi_{b^n}
\end{align*}$$

33
Differentiate $\eta$ with respect to $b^n$ and $d$,
\[ \eta_{b^n} + \eta_T \Phi_{b^n} + \eta_{b^n'} \Phi_{b^n} + \eta_d \Phi_{b^n} + \eta_T \Phi_{b^n}^T = 0 \]
\[ \eta_T \Phi_{b^n}^T + \eta_{b^n'} \Phi_{d^n} + \eta_d \Phi_{d^n} + \eta_T \Phi_{d^n}^T = 0 \]
substitute into $V_{b^n}$ and $V_d$,
\[ V_{b^n} = (1 - \theta)\mu u_{g}^{n} G_{b^n} + \theta v_c C_{b^n} - \lambda \eta_{b^n} \]
\[ V_d = \theta v_c C_d \]
substitute into the last two FOCs,
\[ (1 - \theta)\mu u_{g}^{n} G_{b^n} + \beta \left[ (1 - \theta)\mu u_{g}^{n'} G_{b^n'} + \theta v_c' C_{b^n'} \right] = \lambda \eta_{b^n'} + \beta \lambda \eta_{b^n} \]
\[ \theta v_c C_{d'} + \beta \theta v_c' C_{d} = \lambda \eta_{d'} \]
Expanding the auxiliary functions in (F.23)-(F.24) and (F.27)-(F.28), we get the following Markov optimality conditions.
\[ \{\tau, T^n\} \]
\[ (1 - \theta)[\mu u_{g}^{n} b^n + (1 - \mu) u_{b}^{n} B^n] - \theta v_c [\mu b^n + (1 - \mu) B^n] = \lambda u_{gg} b^n \]
\[ (F.29) \]
\[ \{b^{n'}\} \]
\[ (1 - \theta)[\mu u_{g}^{n} + (1 - \mu) u_{b}^{n}] - \theta v_c [1 + \phi(T^{u} - T)] = \lambda u_{gg} \]
\[ (F.30) \]
\[ \{d'\} \]
\[ \theta v_c S(d') - \beta \theta v_c' = \lambda \Omega_{d'} \]
\[ (F.32) \]
where
\[ \Omega_{x'} = \beta R(b^{n'}) \Phi_{x'}^{n'} \eta_{u_{gg}}^{n'} - \beta R(b^{n'}) (1 - \tau') u_{gg}^{n'} \left[ b^{n'} \Phi_{x'}^{n'} + \Phi_{x'}^{n'} + \Phi_{x'}^{n'} / R(b^{n'}) \right], \quad x = \{b^n, d\} \]
The auxiliary functions are
\[ G_{b^n}^{n} = -(1 - \tau) \]
\[ G_{T}^{n} = b^n \]
\[ G_{T}^{n} = 1 \]
\[ G_{b^n'} = \frac{1}{R(b^{n'})} \]
\[ G_{T}^{n} = B^n \]
\[ G_{T}^{n} = 1 \]
\[ C_{b^n} = -\mu \tau \]
\[ C_{T} = -[\mu b^n + (1 - \mu) B^n] \]
\[ C_{T} = -1 - \phi(T^{u} - T) \]
\[ C_{d} = -1 \]
\[ C_{d'} = \frac{1}{S(d')} \]
\[ \eta_{b^n} = \eta_{gg} G_{b^n}^{n} \]
\[ \eta_{T} = \eta_{gg} G_{T}^{n} \]
\[ \eta_{T} = \eta_{gg} G_{T}^{n} \]
\[
\eta_{b_n} = u_{gg}G_{b_n} + \mu R(b_n')\Phi_{b_n} - \beta R(b_n') (1 - \tau') u_{gg}^* \left[ G_{b_n} + G_{b_n'} \Phi_{b_n} + G_{b_n'} T_{b_n} + G_{b_n'} \Phi_{b_n'} \right]
\]
\[
\eta_{d'} = \beta R(b_n')\Phi_{d} u_{d'}^* - \beta R(b_n') (1 - \tau') u_{gg}^* \left[ G_{d} + G_{d'} \Phi_{d} + G_{d'} T_{d} + G_{d'} \Phi_{d'} \right]
\]

Proofs of propositions in the Markov equilibrium

It is easy to see that Lemma 1 in the paper continues to hold. In addition, following the same steps as in the proof of Lemma E.1, it is straightforward to show \( \lambda \geq 0 \) (we do not even need \( \lambda \) to be strictly positive here). Proposition 2 in the original model becomes:

**Proposition 2a** At the Markov equilibrium steady state, the central government gives the uniform transfer \( T^u < T \) and sets region-specific transfer \( \tau > 0 \).

**Proof.** Define \( \hat{\mu}(b^n, B^o) = \frac{\mu b^n}{\mu b^n + (1 - \mu)B^o} \). We can rewrite (F.29) as

\[
(1 - \theta) \left[ \hat{\mu}(b^n, B^o)u_g^n + (1 - \hat{\mu}(b^n, B^o))u_g^o \right] - \theta v_c = \frac{\hat{\mu}(b^n, B^o)}{\mu} \lambda u_{gg}^n \tag{F.33}
\]

At the steady state,

\[
\lambda u_{gg}^n + \theta v_c \phi(T^u - T) = (1 - \theta) \left[ \mu u_{gg}^n + (1 - \mu)u_g^o \right] - \theta v_c \\
< (1 - \theta) \left[ \hat{\mu}(b^n, B^o)u_g^n + (1 - \hat{\mu}(b^n, B^o))u_g^o \right] - \theta v_c \\
= \frac{\hat{\mu}(b^n, B^o)}{\mu} \lambda u_{gg}^n \\
\leq \lambda u_{gg}^n
\]

The first equality is from (F.30). The first inequality follows because: at the steady state for \( \mu \in (0, 1) \), \( \hat{\mu}(b^n_{ss}, B^o) > \mu \) as \( b^n_{ss} > B^o \geq 0 \), and \( u_g^n > u_g^o \) as \( g^n_{ss} < g^o_{ss} \) (Lemma 1). The second equality follows from (F.33). The second inequality comes from \( \lambda \geq 0 \), \( \hat{\mu} > \mu \) and \( u_{gg}^n < 0 \). Hence, it implies \( \theta v_c \phi(T^u - T) < 0 \), and \( T^u < T \).

To show \( \tau > 0 \) at the steady state, note that \( u_g^o < u_g^n \) (Lemma 1) and (F.30) together imply

\[
(1 - \theta)u_g^o < (1 - \theta) \left[ \mu u_{gg}^n + (1 - \mu)u_g^o \right] = \theta v_c + \lambda u_{gg}^n + \theta v_c \phi(T^u - T) < \theta v_c
\]

as \( T^u < T \), \( \lambda \geq 0 \) and \( u_{gg}^n < 0 \). Since we have established \( T^u < T \), so if \( \tau \leq 0 \), Assumption E.2 would imply \( (1 - \theta)u_g^o > \theta v_c \), which is a contradiction. Therefore, it must be the case that \( \tau > 0 \).

**Remark.** This is consistent with the intuition in the paper. Without the minimum bound on \( T^u \), the Markov central government wants to under-provide the uniform transfer in order to finance a higher \( \tau \), in the presence of heterogeneity in regional debt.

**Proof of Proposition 3 (Over-transfer in Markov)**

**Proof.** We can rewrite (F.30) as

\[
(1 - \theta) \left[ \mu u_{gg}^n + (1 - \mu)u_g^o \right] - \theta v_c = \lambda u_{gg}^n + \theta v_c \phi(T^u - T) < 0
\]

as \( T^u < T \) (Proposition 2a), \( \lambda \geq 0 \) and \( u_{gg}^n < 0 \). Hence,

\[
(1 - \theta) \left[ \mu u_{gg}^n + (1 - \mu)u_g^o \right] < \theta v_c
\]
A three-period model

This section presents a three-period version of the infinite-horizon model in Section 2. We show that the over-transfer result does not hold in period 2 (the final period), though it can hold in the earlier period under some conditions. This is different from the infinite-horizon model, where over-transfer occurs in every period around the steady state.

The timing of the model is as follows. There are three periods: period 1a, period 1b, and period 2. The central and local governments’ actions in each period are listed below:

- **Period 1a.** Local governments start with initial debt $b_1$. The central government starts with revenue $e - f$, chooses the transfer rate $\tau_1$, makes the transfers $\tau_1 b_1$ and spends $c_1 = e - f - \tau_1 b_1$.
- **Period 1b.** Local governments receive own revenues $f$ and transfers $\tau_1 b_1$, choose spending $g_1$ and borrowing $\frac{b_2}{R}$, and repay debt $b_1$.
- **Period 2.** The central government receives revenue $e - f$, chooses its spending $c_2$ and transfers $\tau_2 b_2$. Local governments receive own revenues $f$, transfers $\tau_2 b_2$, repay $b_2$ and spend $g_2$.

For simplicity, we assume the central government does not borrow, and all local governments are identical. We also assume that transfers only consist of distortionary transfers $\tau b$. The interest rate is constant, $R = \frac{1}{\beta}$.

The central government’s inability to commit to transfer policies is reflected in the timing of the model. In this timing, the central government chooses $\tau_1$ after $b_1$ has been given in period 1a. Similarly, $\tau_2$ is chosen in period 2 after $b_2$ is determined in period 1b.

We solve the model backward from period 2.

**Period 2**

**Local Government**

The local government’s problem is trivial in period 2. Because this is the terminal period, no investor is willing to lend and all debt has to be repaid. Therefore, given $b_2$ and $\tau_2$, the local government simply consumes whatever resources it has, $g_2(b_2, \tau_2) = f - b_2 + \tau_2 b_2$.

**Central Government**

For any given $b_2$, the central government in period 2 chooses the transfer rate $\tau_2$ to solve,

$$\max_{\tau_2} \theta v (e - f - \tau_2 b_2) + (1 - \theta) u (f - b_2 + \tau_2 b_2)$$

The first-order condition with respect to $\tau_2$ is

$$\theta v_{c_2} = (1 - \theta) u_{g_2} \quad (G.1)$$

As this is the terminal period, the only effect of the transfer $\tau_2 b_2$ is on closing the central-local spending gap. The central government subsidizes local governments to equalize ($\theta$-weighted) marginal utilities, so there is **no over-transfer in period 2**. Assuming log utility and simplifying, we get

$$\tau_2(b_2) = \frac{(1 - \theta) e - f + \theta b_2}{b_2} \quad (G.2)$$

$$c_2(b_2) = \theta (e - b_2) \quad (G.3)$$

$$g_2(b_2) = (1 - \theta) (e - b_2) \quad (G.4)$$
Period 1b
Local Government
For any given $b_1$ and central government policies $\tau_1$ and $\tau_2$, the local government in period 1b solves the following problem,

$$\max_{b_2} u(b_2) = f + \frac{b_2}{R} - b_1 + \tau_1 b_1 + \beta u(f - b_2 + \tau_2 b_2)$$

The first-order condition with respect to $b_2$ is

$$u_{b_2} = \beta R(1 - \tau_2) u_{b_2}$$

Simplifying,

$$b_2(\tau_1, \tau_2, b_1) = \frac{\tau_2}{(1 + \beta)(1 - \tau_2)} f + \frac{1 - \tau_1}{1 + \beta}$$

$$g_1(\tau_1, \tau_2, b_1) = \frac{1 + \beta - \tau_2}{(1 + \beta)(1 - \tau_2)} f - \frac{1 - \tau_1}{1 + \beta}$$

Note that $\frac{\partial b_2}{\partial \tau_2} > 0$, which implies overborrowing by the local government (relative to the case when there is no distortionary transfer).

What we have described so far in periods 1b and 2 is similar to other two-period or static models in the literature on fiscal decentralization and soft budget constraint. The central government chooses $\tau_2$ after local governments decide $b_2$, so it does not consider the impact of its policy $\tau_2$ on local government’s incentive to borrow. In this case, the local government overborrows and the central government transfers to equalize marginal utilities, i.e., no over-transfer.

Next we consider the impact of $\tau_1$. Substituting $\tau_2$ (equation G.2) into $b_2(\tau_1, \tau_2, b_1)$, we get

$$b_2(\tau_1, b_1) = \frac{\theta b_2(\tau_1, b_1) + (1 - \theta)e - f}{(1 + \beta)[(1 - \theta)b_2 + (1 - \theta)e + f]} f + \frac{1 - \tau_1}{1 + \beta}$$

As $b_2$ shows up on both sides of the equation, equation (G.5) defines $b_2$ as an implicit function of $\tau_1$ for any given $b_1$.

**Lemma G.1.** If $(1 - \theta)e - f > 0$ and $b_1 > 0$, then $\frac{\partial b_2}{\partial \tau_1} < 0$.

**Proof.** Implicitly differentiate (G.5) with respect to $\tau_1$,

$$\frac{\partial b_2}{\partial \tau_1} = \frac{f}{1 + \beta} \frac{f - (1 - \theta)e}{[(1 - \theta)b_2 + (1 - \theta)e + f]^2} \frac{\partial b_2}{\partial \tau_1} - \frac{b_1}{1 + \beta}$$

therefore,

$$\frac{\partial b_2}{\partial \tau_1} = - \frac{[(1 - \theta)b_2 + (1 - \theta)e + f]^2 b_1}{(1 + \beta)[(1 - \theta)b_2 + (1 - \theta)e + f]^2 + f[(1 - \theta)e - f]}$$

$$< 0$$

Lemma G.1 says that, ceteris paribus, transfers in the current period relax local governments’ budget constraint and reduce their (need for) borrowings into the next period.

Period 1a

---

59See Goodspeed et al. (2017) for a review.
Central Government

For any initial local debt position $b_1$ and local government policy function $b_2(\tau_1)$, the central government chooses $\tau_1$ to solve

$$
\max_{\tau_1} \quad \theta v(e - f - \tau_1 b_1) + (1 - \theta) u \left( f + \frac{b_2(\tau_1)}{R} - b_1 + \tau_1 b_1 \right) \\
+ \beta \theta v[\theta(e - b_2(\tau_1))] + \beta (1 - \theta) u [(1 - \theta)(e - b_2(\tau_1))]
$$

Note that we have already substituted equations (G.3) and (G.4) into the objective function. The first-order condition with respect to $\tau_1$ is

$$
0 = \theta \frac{-b_1}{e - f - \tau_1 b_1} + (1 - \theta) \frac{b_1 + \beta \frac{\partial b_2}{\partial \tau_1} + \beta b_2(\tau_1)}{f - b_1 + \tau_1 b_1 + \beta b_2} + \beta \theta \frac{-\theta \frac{\partial b_2}{\partial \tau_1}}{\theta(e - b_2(\tau_1))} + \beta (1 - \theta) \frac{-\theta \frac{\partial b_2}{\partial \tau_1}}{(1 - \theta)(e - b_2(\tau_1))}
$$

(\text{G.6})

**Proposition G.1.** If $(1 - \theta)e - f > 0$, $b_1 > 0$ and $b_2 > 0$, then the central government over-transfers in period 1a, i.e., $(1 - \theta)u_{g1} < \theta v_{c1}$.

**Proof.** We can rewrite equation (G.6) as

$$
\frac{(1 - \theta) b_1}{f - b_1 + \tau_1 b_1 + \beta b_2} - \frac{\theta b_1}{e - f - \tau_1 b_1} = \frac{(1 - \theta) \beta \frac{\partial b_2}{\partial \tau_1} + \beta b_2}{f - b_1 + \tau_1 b_1 + \beta b_2} + \frac{1}{e - b_2}
$$

$$
\Rightarrow \quad b_1 \left[ \frac{1 - \theta}{f - b_1 + \tau_1 b_1 + \beta b_2} - \frac{\theta}{e - f - \tau_1 b_1} \right] = \frac{(1 - \theta) \beta \frac{\partial b_2}{\partial \tau_1} + \beta b_2}{f - b_1 + \tau_1 b_1 + \beta b_2} + \frac{1}{e - b_2}
$$

$$
\Rightarrow \quad b_1 [(1 - \theta)u_{g1} - \theta v_{c1}] = (1 - \theta) \beta \frac{\partial b_2}{\partial \tau_1} (-u_{g1} + u_{g2}) < 0
$$

(\text{G.7})

The inequality follows because $(1 - \theta) \beta \frac{\partial b_2}{\partial \tau_1} < 0$ due to Lemma G.1; and $[-u_{g1} + u_{g2}] > 0$, because local government’s optimality condition is $u_{g1} = \beta R(1 - \tau_2)u_{g2} = (1 - \tau_2)u_{g2} < u_{g2}$, as $\tau_2 > 0$ if $(1 - \theta)e - f > 0$ and $b_2 > 0$ (equation G.2). From equation (G.7), when the local government starts off the period with positive debt ($b_1 > 0$), the central government wants to over-transfer, i.e., $(1 - \theta)u_{g1} < \theta v_{c1}$. \(\square\)

**Discussions.**

Proposition G.1 states that the central government may “over-transfer” in period 1a, consistent with Proposition 3 and the intuition in the paper. In period 1a, the central government is not able to commit to its (future) choice of $\tau_2$, but it understands that a positive $\tau_2$ may induce local governments’ overborrowing (a high $b_2$). Thus in order to reduce the the amount of local borrowing $b_2$, the central government increases its transfers in period 1b ($\tau_1 b_1$), so that local governments’ budget constraint in period 1 is relaxed and the need for a high $b_2$ is lessened.

However, the central government does not over-transfer in period 2, as shown in equation (G.1). This is in contrast to Proposition 3, where the over-transfer property applies in every period. Intuitively, in our three-period model, period 2 is the final period, and the central government’s choice of $\tau_2$ has no impact on the amount of local debt (which has to be zero) by the end of the period. If we extend the model to be more than three (but finite) periods, the over-transfer property will be present in the first period but gradually weaken, until it disappears in the final period.
Using Proposition 2 to simplify the problem, we do not need to solve for the uniform transfer $T^*$. This leaves three policy functions to solve for: $b' = \Phi^b(b^0, d)$, $d' = \Phi^d(b^0, d)$, and $\tau = \Phi^\tau(b^0, d)$.

We solve a system of equations for the policy function and their derivatives with respect to each state. We adopt a local approximation method similar to Klein et al. (2008). We use a linear approximation and assume policy functions are linear functions of states around the steady state. We solve for the three policy functions and their first-order derivatives using a system of nine equations: Local government’s Euler equation (8), central government’s GEEs (D.11) and (D.12), and their derivatives with respect to $b^0$ and $d$

\begin{align*}
0 &= \text{Euler}(b^0, b', \tau, \Phi^b(b', d'), \Phi^\tau(b', d')) \quad \text{(H.1)} \\
0 &= \text{GEE1}(b^0, d, b', \tau, \Phi^b(b', d'), \Phi^d(b', d'), \Phi^\tau(b', d'), \Phi_{b_0}^b, \Phi_{b_0}^\tau) \quad \text{(H.2)} \\
0 &= \text{GEE2}(b^0, d, b', \tau, \Phi^b(b', d'), \Phi^d(b', d'), \Phi^\tau(b', d'), \Phi_{b_0}^b, \Phi_{b_0}^\tau) \quad \text{(H.3)} \\
0 &= d\text{Euler}(b^0, b', \tau, \Phi^b(b', d'), \Phi^\tau(b', d'))/db^0 \quad \text{(H.4)} \\
0 &= d\text{Euler}(b^0, b', \tau, \Phi^b(b', d'), \Phi^\tau(b', d'))/dd \quad \text{(H.5)} \\
0 &= d\text{GEE1}(b^0, d, b', \tau, \Phi^b(b', d'), \Phi^d(b', d'), \Phi^\tau(b', d'), \Phi_{b_0}^b, \Phi_{b_0}^\tau)/db^0 \quad \text{(H.6)} \\
0 &= d\text{GEE1}(b^0, d, b', \tau, \Phi^b(b', d'), \Phi^d(b', d'), \Phi^\tau(b', d'), \Phi_{b_0}^b, \Phi_{b_0}^\tau)/dd \quad \text{(H.7)} \\
0 &= d\text{GEE2}(b^0, d, b', \tau, \Phi^b(b', d'), \Phi^d(b', d'), \Phi^\tau(b', d'), \Phi_{b_0}^b, \Phi_{b_0}^\tau)/db^0 \quad \text{(H.8)} \\
0 &= d\text{GEE2}(b^0, d, b', \tau, \Phi^b(b', d'), \Phi^d(b', d'), \Phi^\tau(b', d'), \Phi_{b_0}^b, \Phi_{b_0}^\tau)/dd \quad \text{(H.9)}
\end{align*}

where $\Phi^b_d$ is the derivative of policy function $\Phi^b$ with respect to $d$, and functional derivatives are computed as follows, for example,

\[
\frac{d\text{Euler}(b^0, b', \tau, \Phi^b(b', d'), \Phi^\tau(b', d'))}{db^0} = \frac{\partial \text{Euler}}{\partial b^0} + \frac{\partial \text{Euler}}{\partial b'} \Phi^b_{b_0} + \frac{\partial \text{Euler}}{\partial \tau} \Phi^\tau_{b_0} \\
+ \frac{\partial \text{Euler}}{\partial b'} (\Phi^b_{b_0} + \Phi^b_{d_0} + \Phi^\tau_{b_0} + \Phi^\tau_{d_0}) + \frac{\partial \text{Euler}}{\partial \tau'} (\Phi^b_{b_0} \Phi^\tau_{b_0} + \Phi^b_{d_0} \Phi^\tau_{d_0})
\]

where $\partial \text{Euler}/\partial b^0$ is approximated by

\[
\frac{\partial \text{Euler}}{\partial b^0} \approx \frac{\text{Euler}(b^0 + \Delta, b', \tau, \Phi^b(b', d'), \Phi^\tau(b', d')) - \text{Euler}(b^0 - \Delta, b', \tau, \Phi^b(b', d'), \Phi^\tau(b', d'))}{2\Delta}
\]

for some tiny value $\Delta$, and second derivatives are set to zero.

\footnote{We checked higher-order (2nd, 3rd) approximations in Mathematica and found that the estimated steady states and first-order derivatives are similar to what we get using a linear approximation.}
I  Additional quantitative results

This section includes additional quantitative results not included in the main text.

I.1  Additional moments for Section 4.1

Table I.1 complements Table 5 and provides additional moments for the baseline counterfactual exercise of Section 4.1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Central-local spending ratio</td>
<td>3.63</td>
<td>1.85</td>
</tr>
<tr>
<td>Local spending</td>
<td>0.062</td>
<td>0.129</td>
</tr>
<tr>
<td>Central spending</td>
<td>0.224</td>
<td>0.239</td>
</tr>
<tr>
<td>Central-to-local transfer</td>
<td>0.039</td>
<td>0.080</td>
</tr>
</tbody>
</table>

Note: Counterfactual experiments use revenue ($e$ and $f$) and spending responsibility share ($\theta$) values from end year, and keep all other parameters at their starting year values. All spending and transfer levels are normalized by national GDP.
### I.2 Quantitative results for Section 4.3.1

Table I.2 shows the calibrated parameter values and the effects of fiscal decentralization on debt levels when governments internalize interest rate schedules.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{B}$</td>
<td>0.0646</td>
<td>0.236</td>
<td>0.237</td>
</tr>
<tr>
<td>$\bar{D}$</td>
<td>0.587</td>
<td>1.063</td>
<td>0.601</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.792</td>
<td>0.654</td>
<td>0.515</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.018</td>
<td>0.018*</td>
<td>0.018*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Local debt</td>
<td>0.062</td>
<td>0.130</td>
<td>0.131 (102%)</td>
<td>0.130</td>
<td>0.122</td>
<td>0.115 (193%)</td>
</tr>
<tr>
<td>Central debt</td>
<td>0.338</td>
<td>0.615</td>
<td>0.380 (15%)</td>
<td>0.615</td>
<td>0.336</td>
<td>0.603 (4%)</td>
</tr>
<tr>
<td>Total government debt</td>
<td>0.400</td>
<td>0.745</td>
<td>0.511 (32%)</td>
<td>0.745</td>
<td>0.458</td>
<td>0.718 (9%)</td>
</tr>
</tbody>
</table>

Note: Assuming governments internalize interest rate schedules and re-calibrating each year. Calibration strategy is identical to the baseline. Total government debt (also known as “general government debt”) is the sum of the local and central government debt. Counterfactual experiments use revenue ($e$ and $f$) and spending responsibility share ($\theta$) values from end year, and keep all other parameters at their starting year values. Numbers in **parentheses** show the percent of debt changes that is explained by decentralization. All debt levels are normalized by national GDP. *We assume $\epsilon$ for 1996 and 2006 is the same as 1988.
I.3 **Quantitative results for Section 4.3.2**

Table I.3 shows the values of internally calibrated parameters with alternative local and central debt interest elasticities.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Target</th>
<th>Calibrated Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{B}$</td>
<td>Local government debt</td>
<td>0.0345</td>
</tr>
<tr>
<td>$\bar{D}$</td>
<td>Central government debt</td>
<td>0.220</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Central-local spending ratio</td>
<td>0.791</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Central-to-local transfer</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Note: Using $\phi_b = 0.05$ and $\phi_d = 0.01$ and re-calibrating each year. Calibration strategy is identical to the baseline. Debt and transfer levels are normalized by national GDP. *We assume $\epsilon$ for 1996 and 2006 is the same as 1988.
I.4 Quantitative results for Section 4.3.3

Table I.4 shows the calibrated parameter values and the effects of fiscal decentralization on debt levels with alternative calibration of $\mu$ and $B^o$.

**Table I.4: Counterfactual Experiment with alternative calibration of $\mu$ and $B^o$**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{B}$</td>
<td>0.00358</td>
<td>0.137</td>
<td>0.139</td>
</tr>
<tr>
<td>$\bar{D}$</td>
<td>0.316</td>
<td>0.594</td>
<td>0.319</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.788</td>
<td>0.652</td>
<td>0.514</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.018</td>
<td>0.018$^*$</td>
<td>0.018$^*$</td>
</tr>
</tbody>
</table>

Moments

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Local debt</td>
<td>0.062</td>
<td>0.130</td>
<td>0.147</td>
<td>0.122</td>
</tr>
<tr>
<td>Central debt</td>
<td>0.338</td>
<td>0.615</td>
<td>0.389</td>
<td>0.615</td>
</tr>
<tr>
<td>Total government debt</td>
<td>0.400</td>
<td>0.745</td>
<td>0.536</td>
<td>0.745</td>
</tr>
</tbody>
</table>

Note: Using $\mu = 0.6$ and $B^o = 0.046$ and re-calibrating each year. Calibration strategy is identical to the baseline. Total government debt (also known as “general government debt”) is the sum of the local and central government debt. Counterfactual experiments use revenue ($\epsilon$ and $f$) and spending responsibility share ($\theta$) values from end year, and keep all other parameters at their starting year values. Numbers in parentheses show the percent of debt changes that is explained by decentralization. All debt levels are normalized by national GDP. *We assume $\epsilon$ for 1996 and 2006 is the same as 1988.
I.5 Quantitative results for Section 4.3.4

Table I.5 shows the effects of fiscal decentralization on debt levels with an alternative value of \( \epsilon \).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Local debt</td>
<td>0.149</td>
<td>0.284</td>
</tr>
<tr>
<td>Central debt</td>
<td>0.482</td>
<td>1.029</td>
</tr>
<tr>
<td>Total government debt</td>
<td>0.631</td>
<td>1.313</td>
</tr>
</tbody>
</table>

Note: Using \( \epsilon = 0.1 \), and all other parameters identical to the baseline. Moments are not matched to data. Total government debt (also known as “general government debt”) is the sum of the local and central government debt. Counterfactual experiments use revenue (\( e \) and \( f \)) and spending responsibility share (\( \theta \)) values from end year, and keep all other parameters at their starting year values. Numbers in parentheses show the percent of debt changes that is explained by decentralization. All debt levels are normalized by national GDP.
I.6 Quantitative results for Section 4.3.5

Table I.6: Counterfactual Experiment using moments from simulated paths

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Calibrated Values in simulated 1996 economy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>0.594</td>
</tr>
<tr>
<td>$D$</td>
<td>1.080</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.586</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.014*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1988–1996 (1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steady state 1988</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local debt</td>
<td>0.062</td>
<td>0.130</td>
<td>0.103 (59%)</td>
</tr>
<tr>
<td>Central debt</td>
<td>0.337</td>
<td>0.613</td>
<td>0.359 (8%)</td>
</tr>
<tr>
<td>Total government debt</td>
<td>0.399</td>
<td>0.743</td>
<td>0.462 (18%)</td>
</tr>
</tbody>
</table>

Note: The calibration finds the set of parameters ($B, D, \theta, \epsilon$) such that the moments—total local government debt, central government debt, central-local spending ratio, and central-to-local transfer—in the simulated 1996 economy (off-steady state) match those in the data for 1996. Total government debt (also known as “general government debt”) is the sum of the local and central government debt. Counterfactual experiments use revenue ($\epsilon$ and $f$) and spending responsibility share ($\theta$) values from 1996, and keep all other parameters at the 1988 values. Numbers in parentheses show the percent of debt changes that is explained by decentralization. All debt levels are normalized by national GDP. *We assume $\epsilon$ for 1996 is the same as 1988.

Figure I.1: Simulated path of total debt from 1998 steady state

Note: The plot shows the path of total debt simulated from the 1988 steady state economy. The parameters of the economy are changed to calibrated 1996 values from 1989 onward. For the counterfactual economy, we only change the fiscal decentralization parameters ($\epsilon, f, \theta$) and keep the others at their 1988 levels.

REFERENCES


——— (2005b): “And the last shall be first: federalism and soft budget constraints in Germany,” *Unpublished paper. Department of Political Science, Massachusetts Institute of Technology*.


