Journal of Monetary Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jmoneco

Dynamic tax externalities and the U.S. fiscal transformation☆

Martín Gonzalez-Eiras[a], Dirk Niepelt[b,c,*]

[a] University of Copenhagen, Oster Farimagsgade 5, Copenhagen K 1353, Denmark
[b] Study Center Gerzensee, P.O. Box 21, Gerzensee 3115, Switzerland
[c] University of Bern, Schanzeneckstrasse 1, Bern 3001, Switzerland

ARTICLE INFO

Article history:
Received 4 June 2018
Revised 15 February 2019
Accepted 27 February 2019
Available online xxx

Keywords:
Fiscal federalism
Tax centralization
Grants
Politico-economic equilibrium

ABSTRACT

We propose a theory of tax centralization in politico-economic equilibrium. Taxation has dynamic general equilibrium implications which are internalized at the federal, but not at the regional level. The political support for taxation therefore differs across levels of government. Complementarities on the spending side decouple the equilibrium composition of spending and taxation and create a role for inter governmental grants. The model provides an explanation for the centralization of revenue, introduction of grants, and expansion of federal income taxation in the U.S. around the time of the New Deal. Quantitatively, it accounts for approximately 30% of the federal revenue share's doubling in the 1930s, and for the long-term increase in federal grants.

1. Introduction

Whether control over fiscal policy should rest with national, regional or local governments depends on how effective these agents make use of their authority. When information frictions render it difficult to cater to heterogeneous needs, fiscal policy is best chosen de-centrally. When it is key to internalize spillover effects, in contrast, centralized policy choices are advantageous. A broad body of fiscal federalism literature has studied the normative and positive implications of this fundamental trade-off. The focus of that literature has generally been on static sources of the cost-benefit differences across governments.

In this paper, we propose a complementary—dynamic—source that arises from general equilibrium effects of taxation. We adopt a positive perspective and show that this dynamic source of cost differences across governments is present under a wide set of assumptions. Moreover, we argue that it provides a novel explanation of the dramatic fiscal transformation in the United States during the 1930s when federal tax collections increased strongly and federal grants started their long-term rise.

The model features overlapping generations that work, save, consume, and vote as well as a central, or federal, government and many regional governments that impose labor income taxes to finance the provision of public services.[1] In

☆ For useful comments and discussions, we thank the editor and two anonymous referees as well as Asger Lau Andersen, Maximilian von Ehrlich, David Dreyer Lassen and seminar participants at CESifo macro meeting, CEMFI, IIBEO workshop, SED, Universidad Carlos III, University of Copenhagen, VfS macro meeting.

* Corresponding author at: Study Center Gerzensee. P.O. Box 21, 3115 Gerzensee, Switzerland.
E-mail addresses: mge@alum.mit.edu (M. Gonzalez-Eiras), dirk.niepelt@szgerzensee.ch (D. Niepelt).
URL: http://www.alum.mit.edu/www/mge (M. Gonzalez-Eiras), http://www.niepelt.ch (D. Niepelt)

[1] We refer to a state with a multi-tier political organization as a "federal" state, and to a government that makes decisions at the central level as a "federal" government. We refer to governments making decisions at the local level as "regional" governments.

https://doi.org/10.1016/j.jmoneco.2019.02.007
0304-3932/

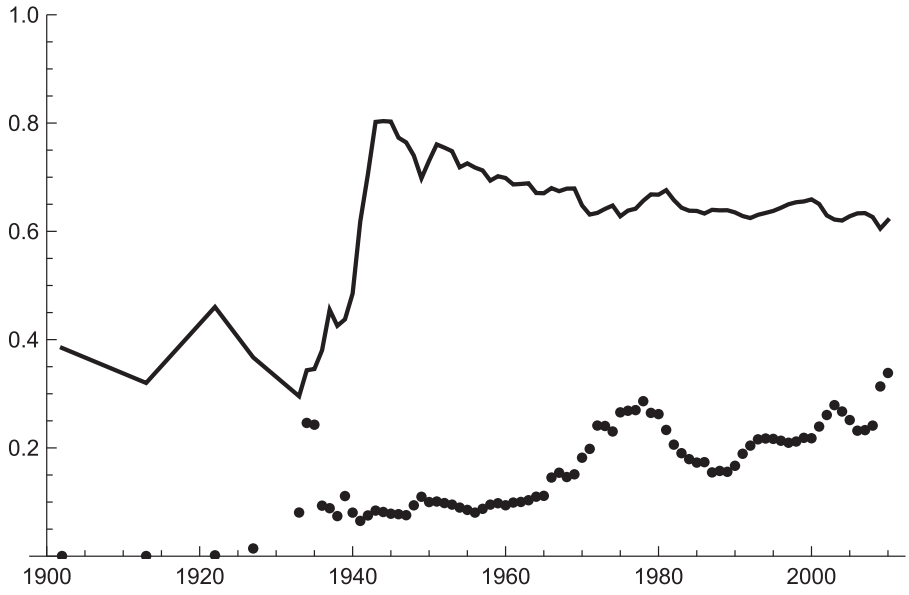

Fig. 1. Fiscal transformation in the United States: Federal revenues and grants. Federal relative to total government revenues (solid), and federal grants relative to state and local revenues (dots). Sources: Wallis (2000) for years 1902, 1913, 1922, 1927; NIPA tables for subsequent years.

politico-economic equilibrium, households make optimal savings choices conditional on current and expected future policies; and they vote for their preferred political candidates, taking the policy functions of other political decision makers as well as the competitive equilibrium conditions into account.

Taxation slows down capital accumulation and thus has general equilibrium effects: It drives up interest rates and lowers future wages. Voters and policy makers at the federal level—rationally—internalize these general equilibrium effects to the extent that they are affected by them.[2] In contrast, voters and policy makers at the regional level—rationally—do not perceive general equilibrium effects of their decisions since regions are small relative to the nation and markets are not segmented. As a consequence, the net cost of a federal tax hike as perceived by a voter participating in national elections differs from the net cost of a regional tax hike as perceived in regional elections.[3]

In the baseline specification, federal and regional spending are perfect substitutes and all static sources of cost-benefit differences across governments are absent: Government spending does not generate externalities; preferences for public services are uniform across the population; and regional tax bases are immobile, undermining any motive for tax competition. We show that, nevertheless, the incentives to raise taxes at the federal level differ from those at the regional level. Depending on the sign of the dynamic general equilibrium effects, federal taxation enjoys stronger or weaker support in politico-economic equilibrium than regional taxation, and the equilibrium composition of tax collections across governments is determinate.

This result is robust along many dimensions. We introduce labor mobility across regions and find that this does not fundamentally alter our findings. We allow for elastic labor supply, tax distortions, and additional policy instruments and show that the results are robust since the perceived cost differences due to general equilibrium effects are orthogonal to the effects of tax distortions. We also find that our results are robust to introducing policy instruments for intergenerational redistribution, such as public debt or social security, or longer-lived households.

We also consider the effects of capital income taxation. In contrast to labor income taxes which depress workers' savings, capital income taxes do not affect future capital accumulation because they are chosen ex post and reduce the income of the old.[4] From the perspective of federal and regional voters, the net cost of taxation thus is the same and the asymmetry in the political support for federal versus regional taxation disappears. A shift from capital to labor income taxation therefore can trigger a major change in the composition of tax collections, favoring federal or regional income taxation.

We argue that this mechanism offers a novel explanation for the dramatic fiscal transformation that the United States underwent during the 1930s, see Figs. 1 and 2.[5] On the eve of the Great Depression, local governments collected the ma-

[2] The welfare consequences for yet unborn cohorts who are not represented in the political process are not internalized.

[3] Empirical evidence suggests that voters understand which level of government is responsible for policy decisions of relevance for them. See, for example, Hilt and Rahn (2018) on the effect of price depreciation of "liberty bonds" in the interwar period on voting outcomes in Congressional and Presidential elections.

[4] When cohorts live longer than for two periods then capital income taxes do affect capital accumulation but by less than labor income taxes. The perceived difference in tax collection costs across governments then is smaller than with labor income taxes.

[5] See Wallis (2000) for a discussion of this and two earlier transformations of the American fiscal architecture.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

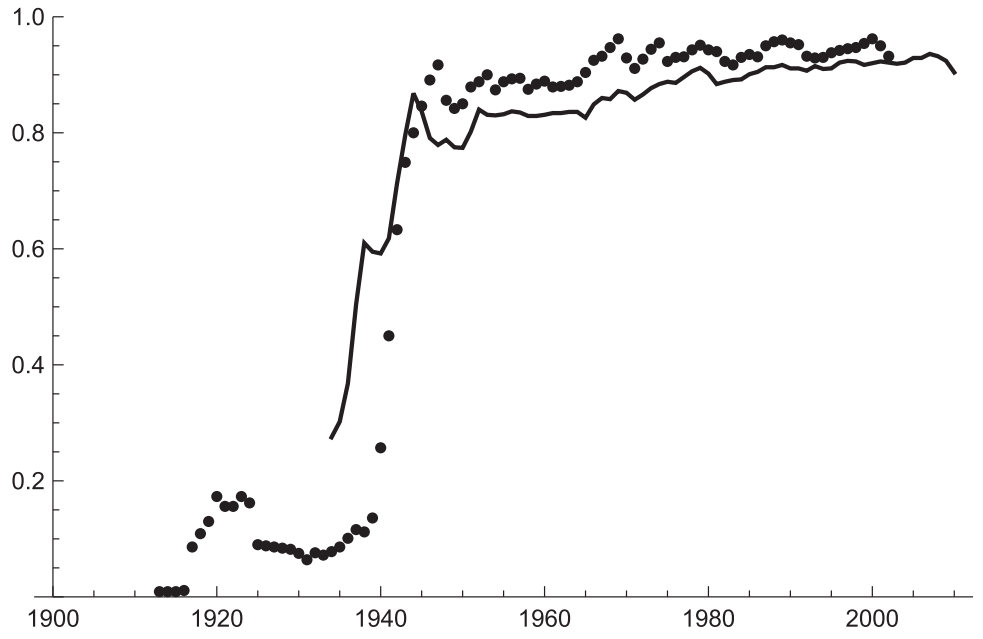

Fig. 2. Fiscal transformation in the United States: Federal income taxation. Number of tax returns relative to number of tax units (dots), and federal income tax (including OASDI) relative to total revenue of the federal government (solid). Sources: Piketty and Saez (2003), Table A0, and Office of Management and Budget, Fiscal Year 2016, Historical Tables, Table 2.2.

jority of tax revenues and property taxes accounted for nearly half of all revenues. The federal government's main source of revenue were tariffs, and on a smaller scale, property taxes. In the 1930s this arrangement changed completely.[6] As indicated by the solid line in Fig. 1 the revenue share of the federal government nearly doubled and inter governmental grants (indicated by dots and discussed below) emerged as a central source of revenue for state and local governments. An even more dramatic transformation occurred with respect to the federal tax base, see Fig. 2. The income tax share of federal revenues more than doubled within a few years and continued to grow quickly for another decade, and the share of tax units who paid federal income tax similarly exploded.

Our model explains this transformation as the equilibrium response to the ratification of the Sixteenth Constitutional Amendment which introduced the *possibility* for the federal government to tax income.[7] Starting from a situation with exclusive competence for income taxation at the level of the states and strong reliance on property taxes (which generate limited general equilibrium effects), the ratification opened the door for the federal government to tax labor income and to exploit the general equilibrium effects we emphasize.[8] Stronger demand for government outlays, specifically for New Deal policies and spending related to World War II, subsequently strengthened the *incentives* to employ the newly available tax instrument, and the fiscal transformation took place.

Importantly, our model does not aim at explaining the increase in *total* government revenues. This increase occurred against the background of the New Deal, World War II, and the associated macroeconomic developments and, as we discuss below, has been rationalized based on theories of the role of war, patriotism, and the media for tax policy.[9] Nor, therefore, does our model aim at capturing the within-cohort distributive conflicts as well as administrative and legal hurdles,[10] which had to be overcome before the federal government could start levying a comprehensive income tax, and which were brushed

[6] See Wallis and Oates (1998) for a description of New Deal programs and a discussion of federal deficits that accompanied the transformation.

[7] The Sixteenth Amendment which was passed in 1913 and implemented in the course of the following years, states: "The Congress shall have power to lay and collect taxes on incomes, from whatever source derived, without apportionment among the several States, and without regard to any census or enumeration."

[8] Income taxes predominantly affect savers. While they are collected from both workers (who save) and retirees (who do not), the fraction of individuals paying taxes sharply falls with age, see for example Greenstone and Looney (2012). Applying Piketty and Saez's 2003 methodology to classify respondents we find that in the 2015 March Current Population Survey 87% of tax units aged 65 or below paid taxes, roughly twice the share of those aged 65 years or older.

[9] See, for example, Besley and Persson (2009), Vélez (2014), and Jones (1988). Our model also does not aim at explaining regional funding problems, for instance due to temporary changes in property prices or public assistance needs.

[10] The Treasury underwent a major reorganization; the number of employees at the Bureau of Internal Revenue increased fourfold; and the Supreme Court upheld the constitutionality of tax legislation enacted based on the Amendment (in 1916) and made a series of decisions relating to the proper definition of income and the fairness of its taxation (after the First World War), see Mehrotra (2013). All of this happened against the backdrop of political conflict as to who should pay income tax and how progressive the system should be, reflected in sharp swings in the highest marginal tax rate and the share of the population that was tax liable.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

aside by the need to increase expenditure for New Deal policies and World War II spending.[11] Instead, our model aims at explaining the shift in the *composition* of government financing, namely the centralization of revenue collection, as well as the rise of federal grants that took place in parallel to the aggregate revenue increase.[12]

To study the latter feature of the fiscal transformation—the rise of federal grants—we introduce a motive to decouple revenue collection from government spending. This requires the composition of both taxes and spending across governments to be determinate. We therefore relax the assumption that regional and federal spending are perfect substitutes and assume instead that complementarities between the two spending components exist. The grant instrument has value when it is beneficial to channel revenue from the federal government to regions because tax revenue at the federal level is "cheap." When we additionally introduce static cross-regional externalities, a key element in many analyses of fiscal federalism, then channeling revenue to regions also has value when the latter spend too little from the national perspective.

The extended framework with complementarities, static cross-regional externalities and another common element of fiscal federalism models, namely regional preference heterogeneity, is as analytically tractable as the baseline model. In line with empirical evidence, it predicts grants to crowd out local taxation.[13] Quantitatively, the calibrated model is able to explain the trend increase of grants as the result of time varying preference heterogeneity regarding the size of government; based on empirical evidence, we associate this heterogeneity with disparities between urban and rural regions. The model's explanatory power for the federal tax share also is substantial; the extended framework explains roughly 30% of the observed increase between 1930 and 1950.[14] We find that alternative explanations for the fiscal transformation that rely on static spending externalities as they are discussed in the fiscal federalism literature would require unreasonably large increases in static externalities to generate similar quantitative results.

Related Literature. We build on the classic analysis of fiscal federalism that stresses trade-offs between forces favoring centralization and decentralization. Oates (1972) finds that absent spillovers and cost-savings from centralized tax collection or public good provision, heterogeneous preferences render decentralization preferable. Without information frictions, a centralized system may in principle support differentiated provision (Oates, 1999). But various political economy frictions favor uniform centralized policy choices.[15] Alesina and Spolaore (1997) analyze the effect of international integration on the costs and benefits of centralization and thus, the number of countries.

Similar arguments are discussed in the political science literature (e.g., Kincaid, 2011) which tends to favor federalist governance structures for diverse countries. Treisman (2007) questions many rationales for and against political decentralization. He argues that administrative efficiency requires administrative, not political decentralization and he criticizes the view that local governments better manage local information.[16] Our argument is related in so far as it stresses the decoupling of tax and spending decisions.

Horizontal and vertical tax competition in federal structures gives rise to important static externalities. Gordon (1983) shows that uncoordinated taxation of mobile factors gives rise to revenue (and other) externalities across regions. A federal government concerned with welfare at the national level may correct some of these externalities by imposing federal taxes or extending federal grants. Our paper also emphasizes tax externalities but of a different type, namely dynamic externalities due to general equilibrium effects, and it builds a positive theory of fiscal federalism and federal grants.[17]

Uniform federal grants combined with non-uniform federal taxes (or vice versa) redistribute between regions and may constitute a form of inter-regional risk sharing (see, for example, Persson and Tabellini, 1996). The fact that such risk-sharing is very common does not provide a rationale for federal grants, however, since risk sharing in the joint interest of regions can be implemented without federal intervention. In our model, fiscal policy does not redistribute, and grants are used to achieve an allocation of resources that regions would not choose by themselves.

Wallis (2000) documents that the U.S. passed through distinct regimes of government finance and suggests that the costs of raising revenue differ across governments. Our model provides an explanation for such cost differences that stresses gen-

[11] Voters' behavioral biases, which we do not model, might have constituted an additional source of delay. In particular, voters might initially have lacked the understanding of general equilibrium effects of federal income taxation; see Dal Bó et al. (2018) and Agranov and Palfrey (2015) for evidence from lab experiments on under appreciation of general equilibrium effects in the short run that disappears with experience.

[12] Note that, conceptually, changes in the composition of government financing and intergovernmental grants are unrelated to the total size of government. They are also unrelated to funding problems of specific governments since not only funds, but also spending responsibilities can be reallocated.

[13] For example, Knight (2002) finds statistically and economically significant crowding out for the Federal Highway Aid Program in the U.S. He addresses identification problems (an omitted variable bias due to the positive correlation between grant levels and unobserved preferences for public spending) by using the political power of state congressional delegations as instruments.

[14] A calibration based on the baseline model explains all the observed increase in the federal tax share.

[15] For example, legislative bargaining among regional representatives at the federal level may imply reduced sensitivity of policy to regional needs (Lockwood, 2002); differentiated central service provision can give rise to costly bargaining and delay and may thus be avoided (Harstad, 2007); credibility problems in signalling local tastes to the central government may generate inefficient federal policy choices (Kessler, 2014); and centralization to increase accountability may have to be accompanied by policy uniformity because otherwise, the central government would implement policies favoring regions that monitor more extensively (Boffa et al., 2016).

[16] According to Treisman (2007), decentralization is important for policy stability and centralization is important for fiscal coordination.

[17] Hatfield and Padró i Miquel (2012) study an economy where some public goods are funded and provided regionally and others federally. They show that the federal government imposes capital income taxes while regions resort to lump sum taxes, due to tax competition. In our setting, the grant instrument decouples funding from public good provision.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

eral equilibrium effects.[18] This explanation complements alternative, static theories that rely on permanently lower information processing costs for the federal government;[19] permanently higher externalities from public infrastructure investment; or interstate mobility and tax competition which have been criticized.[20] In addition to offering a novel source of differences in the cost of taxation our model can quantitatively account for important parts of the dramatic transformation of the U.S. fiscal system during the 1930s and afterwards.

On the methodological side, our paper relates to the literature on dynamic politico-economic equilibrium (Krusell et al., 1997). While most work in this literature studies equilibria with a single political decision maker, Song et al. (2012) analyze politico-economic equilibrium in a setting with a continuum of governments that take factor prices as given. We solve a dynamic game with a continuum of regional governments and a central government that internalizes general equilibrium effects.

Outline. The remainder of the paper is structured as follows. In Section 2 we describe the model, and in Section 3 we define equilibrium. Sections 4 and 5 contain the analysis of the baseline model and its extension. In Section 6, we contrast the model's quantitative implications with empirical evidence on the fiscal transformation in the U.S. during the 1930s and subsequent fiscal trends. Section 7 concludes. The online appendix contains proofs and ancillary discussions.

2. The model

2.1. Demographics and institutions

We consider an economy inhabited by overlapping generations of workers and retirees. Workers supply labor, pay taxes, consume and save. In the subsequent period, they retire, consume the return on their savings, and die. The ratio of workers to retirees in period t equals ν_t and follows a deterministic process.

The economy is composed of a continuum of regions of measure one over the unit interval. Each region is populated by a continuum of homogeneous agents. The population structure including the preferences of agents is the same across regions (we relax this assumption later). Regions are indexed by i.

Policy decisions are taken by governments at the federal and the regional level. Federal and regional governments act in the interest of voters participating in nationwide and regional elections, respectively. None of the governments can commit, and in each period they take decisions simultaneously.[21]

2.2. Production of final good

A continuum of competitive firms transforms capital and labor into output. Capital is owned by retirees—it corresponds to the savings of workers in the preceding period—and fully depreciates after a period. The economy-wide capital stock per worker, k_t, therefore corresponds to the economy-wide per-capita savings of workers in the previous period, s_{t-1}, normalized by ν_t. Labor is supplied inelastically (we relax this assumption later). The gross interest rate R_t and the wage w_t are determined competitively.

We assume that the production function displays constant returns to scale such that factor prices in period t only depend on k_t,

$$R_t = R(k_t), \;\; w_t = w(k_t). \tag{1}$$

Moreover, we assume that the elasticities of the factor prices with respect to the capital-labor ratio, ϵ_{Rk} and ϵ_{wk}, are independent of the latter, $\epsilon_{Rk}, \epsilon_{wk} \perp k_t$. Examples of production functions that satisfy these assumptions include the Cobb-Douglas production function with capital share α where factor prices equal $R_t = \alpha k_t^{\alpha-1}$ and $w_t = (1-\alpha)k_t^{\alpha}$, the Ak production function, or a small open economy with exogenous factor prices.[22]

2.3. Production and financing of public services

The quantity or quality of publicly provided services (or public services, for short) in a region i, g_t^i, depends on public spending at the regional level and nationwide. Let e_t^i denote spending at the regional level and e_t the—uniform—spending

[18] In other contexts, Kotlikoff and Rosenthal (1990), Soares (2005), and Gonzalez-Eiras and Niepelt (2008) have emphasized the role of general equilibrium effects in politico-economic equilibrium.

[19] Wallis (2000) suggests that the introduction of Social Security payroll taxes could have lowered federal tax collection costs.

[20] Rhode and Strumpf (2003) document that households' migration decisions mostly reflect personal factors rather than Tiebout (1956) sorting and they find that the secular decrease in mobility costs in the United States was not accompanied by stronger policy or preference heterogeneity across communities. Rhode and Strumpf (2003) conclude that "any theoretical or empirical model that adopts a pure Tiebout framework...is misspecified" (p. 160). Similarly, Baicker et al. (2012) find that "patterns in mobility, seem to have little power to explain observed changes in the landscape of fiscal federalism" (p. 1080).

[21] In the data, this is not strictly true as state and federal elections of the executive branches are not perfectly synchronized. Our choice of timing assumption is motivated by our interest in the long run determinants of fiscal federalism and the fact that one period in the model corresponds to several decades.

[22] The independence assumption can be disposed of at the cost of losing the ability to derive closed-form solutions.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

by the federal government.[23] In the baseline model, we assume that federal and regional spending are perfect substitutes in the production of public services and we abstract from externalities across regions (we relax both assumptions below). Accordingly,

$$g_t^i = e_t^i + e_t. \tag{2}$$

Spending by the federal government is financed by a labor income tax at rate τ_t and spending by region i is financed by a tax at rate τ_t^i. (Below, we introduce federal grants as an additional source of regional revenue.) Since all governments balance their budget in each period this implies

$$e_t = w_t \tau_t, \quad e_t^i = w_t \tau_t^i. \tag{3}$$

Tax rates are non-negative.

2.4. Preferences and household choices

Workers and retirees in region i and period t value private consumption, $c_{1,t}^i$ and $c_{2,t}^i$ respectively, as well as public services. Workers discount the future at factor $\beta \in (0, 1)$. For analytical tractability, we assume that period utility functions are logarithmic. Welfare of a worker who chooses savings s_t^i is given by

$$\ln(c_{1,t}^i) + \gamma_t \ln(g_t^i) + \beta\left(\ln(c_{2,t+1}^i) + \gamma_{t+1} \ln(g_{t+1}^i)\right)$$
$$\text{s.t.} \quad c_{1,t}^i = w_t(1 - \tau_t - \tau_t^i) - s_t^i, \quad c_{2,t+1}^i = s_t^i R_{t+1}.$$

Parameter γ_t represents the preference for public services.

Taking prices and taxes as given the worker optimally chooses

$$s_t^i = \frac{\beta}{1+\beta} w_t(1 - \tau_t - \tau_t^i), \tag{4}$$

that is, equilibrium consumption and saving of a worker are proportional to the after tax wage. Accordingly, lifetime utility of a worker is proportional to the log after tax wage, the log gross interest rate, and the logarithms of current and future public services. We summarize this information in the indirect utility function which is given by (dropping irrelevant constants)

$$U_t^{i,w} = (1+\beta)(\ln(w_t) + \ln(1 - \tau_t - \tau_t^i)) + \beta \ln(R_{t+1}) + \gamma_t \ln(g_t^i) + \beta \gamma_{t+1} \ln(g_{t+1}^i) \tag{5}$$

subject to (2), (3). Similarly, the indirect utility function of a retiree with personal savings s_{t-1}^i equals

$$U_t^{i,r} = \ln(s_{t-1}^i) + \ln(R_t) + \gamma_t \ln(g_t^i) \tag{6}$$

subject to (2), (3).

2.5. Elections

Elections take place at the beginning of each period, simultaneously in all regions and nationwide. We assume that preferences are aggregated through probabilistic voting.[24] At each level of government policy thus maximizes a convex combination of the objective functions of all groups of voters, where the weights reflect the groups' sizes and their responsiveness to policy changes. We assume that across regions, voters are equally responsive to proposed changes in policy platforms. However, we allow for age related variation in responsiveness, reflected in a per capita political influence weight of unity for young voters and a per capita weight of $\omega \geq 0$ for retired voters.

3. Equilibrium

3.1. Competitive equilibrium

The state is given by z_t, which includes the exogenous demographic parameter as well as the cross section of savings levels across regions which we denote by $\vec{s}_{t-1}$. (Throughout the paper, we indicate cross sections by an arrow.) Conditional on z_t, the production function as well as competition among firms determine factor prices, w_t and R_t. A financing policy (or policy for short) of all regions and the federal government, $(\vec{\tau}_t, \tau_t)$, then determines public services, $\vec{g}_t$, capital accumulation, $\vec{s}_t$, and thus z_{t+1}. Conditional on z_t, a policy sequence $\{\vec{\tau}_s, \tau_s\}_{s \geq t}$ thus fully determines an allocation and price system.

We focus on symmetric equilibria where all regions behave identically, except possibly a set of regions of measure zero. We denote the "typical" regional tax by τ_t^j, and "typical" public services by g_t^j.

Definition 1. A *competitive equilibrium* conditional on z_0 and a policy sequence $\{\tau_t^j, \tau_t\}_{t \geq 0}$ is given by an allocation and price system such that

[23] Thus, we allow for both levels of government to tax and spend. For rationalizations of policy uniformity at the federal level, see the literature review in the introduction.

[24] See Lindbeck and Weibull (1987). In the online appendix we offer a formal discussion of probabilistic voting.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

1. Capital evolves according to $k_t = s_{t-1}/\nu_t$, and factor prices are determined according to (1) for all t;
2. the government budget constraints (2) and (3) are satisfied for all t; and
3. households optimize: (4) is satisfied for all i, t.

3.2. Politico-economic equilibrium

In politico-economic equilibrium political decision makers optimally choose the values of the policy instruments under their control, taking all implications of their actions into account and forming rational expectations about future policy choices. We assume that these choices are Markov that is, they are functions of the fundamental state variables. We conjecture—and later verify—that policy choices are independent of the endogenous state variables, $\vec{s}_{t-1}$, such that future policy choices are unaffected by current policy choices.[25]

Political decision makers at the regional and federal level perceive the economic environment differently. On the regional level they take policy choices by the federal government and in other regions, as well as factor prices and externalities, as given. On the federal level they take regional policy choices as given and account for the endogeneity of factor prices.

Formally, under the conjecture a regional decision maker at date t takes $(w_t, w_{t+1}, R_t, R_{t+1})$ as well as s^i_{t-1} and $(\tau^j_t, \tau_t, \tau^i_{t+1}, \tau^j_{t+1}, \tau_{t+1})$ as given and her objective is $\omega U^{i,r}_t/\nu_t + U^{i,w}_t$. That is, the regional decision maker maximizes

$$V^i_t \equiv \left(\frac{\omega}{\nu_t} + 1\right)\gamma_t \ln(g^i_t) + (1+\beta)\ln(1-\tau_t-\tau^i_t) \quad \text{s.t.} \quad (2), (3) \tag{7}$$

In contrast, the federal decision maker at date t takes (w_t, R_t) as well as s_{t-1} and $(\tau^j_t, \tau^j_{t+1}, \tau_{t+1})$ as given and is concerned with $\omega U^{j,r}_t/\nu_t + U^{j,w}_t$. Unlike its regional counterpart, the federal decision maker internalizes the effects of policy on w_{t+1} and R_{t+1} and thus, on the provision of public goods and consumption of private goods for current workers next period. The federal decision maker thus maximizes

$$V_t \equiv \left(\frac{\omega}{\nu_t} + 1\right)\gamma_t \ln(g^j_t) + (1+\beta)\ln(1-\tau_t-\tau^j_t) + \beta\ln(R_{t+1}) + \beta\gamma_{t+1}\ln(g^j_{t+1}) \tag{8}$$
$$\text{s.t.} \quad (1), (2), (3), (4), k_{t+1} = s_t/\nu_{t+1}.$$

We can now define politico-economic equilibrium (under the conjecture).[26]

Definition 2. A *politico-economic equilibrium* conditional on z_0 is given by a policy sequence $\{\tau^j_t, \tau_t\}_{t\geq 0}$ and an allocation and price system such that

1. $\tau^i_t \geq 0$ maximizes V^i_t and $\tau^i_t = \tau^j_t$ for all i, t;
2. $\tau_t \geq 0$ maximizes V_t for all t; and
3. the allocation and price system constitute a competitive equilibrium conditional on z_0 and $\{\tau^j_t, \tau_t\}_{t\geq 0}$.

The politico-economic equilibrium characterized in Definition 2 is the unique symmetric Markov perfect equilibrium arising in the limit of the finite-horizon economy.[27]

4. Tax centralization

Absent heterogeneity in regional preferences or static externalities from spending or taxation across regions, none of the traditional static fiscal federalism motives for decentralization or centralization is present. Nevertheless, the equilibrium degree of centralization of tax collections generally *is* determinate. To see this, consider the derivative of the regional objective function V^i_t with respect to the regional tax rate, τ^i_t (which equals τ^j_t in equilibrium), and the derivative of the federal objective function V_t with respect to the federal tax rate, τ_t. Since tax rates must be non-negative the derivatives of V^i_t in (7) and of V_t in (8) must be weakly negative in equilibrium,

$$\left(\frac{\omega}{\nu_t} + 1\right)\frac{\gamma_t}{\tau^j_t + \tau_t} - \frac{1+\beta}{1-\tau_t-\tau^j_t} \leq 0, \tag{9}$$

$$\left(\frac{\omega}{\nu_t} + 1\right)\frac{\gamma_t}{\tau^j_t + \tau_t} - \frac{1+\beta}{1-\tau_t-\tau^j_t} + \mathcal{F}_t \leq 0, \tag{10}$$

[25] This conjecture is motivated by two observations. First, the indirect utility functions are additively separable in prices and policy (reflecting our assumption about preferences and the production of public services); and second, the elasticities of factor prices with respect to the capital-labor ratio are orthogonal to the latter (reflecting our assumption about the aggregate production function).

[26] In general, politico-economic equilibrium requires that political decision makers anticipate future policy choices to be determined according to policy functions (mappings from the state into policy) and that optimal policy choices are consistent with policy functions evaluated at the state. Under the conjecture this consistency requirement is trivially satisfied.

[27] This follows from a simple backward induction argument. In the last period, a unique policy combination constitutes the Nash equilibrium of the static game played by the federal and regional decision makers. Anticipating this outcome, the equilibrium policy choices of players in the last period but one are unique as well, et cetera. See, for example, Gonzalez-Eiras and Niepelt (2008).

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

where $\mathcal{F}_t \equiv -\beta(\epsilon_{Rk} + \epsilon_{wk}\gamma_{t+1})/(1 - \tau_t - \tau_t^j)$. In addition, the corresponding complementary slackness conditions must be satisfied.

The terms in the first inequality represent the marginal benefit and cost, respectively, of a higher regional tax rate as perceived by voters at the regional level. The marginal benefit derives from higher public services which both old and young voters appreciate, and the marginal cost reflects reduced wealth and thus, consumption of workers. In the second inequality, the first two terms represent the marginal benefit of higher public services and the direct marginal cost of lower consumption as perceived by voters in nationwide elections. The marginal benefit and the direct marginal cost are the same as those perceived on the regional level because of the uniformity of preferences and the absence of static spending externalities.

The third term in the second inequality, $\mathcal{F}_t$, represents the *indirect* net benefit of higher taxes due to general equilibrium factor price effects that young voters at nationwide elections internalize. This net benefit materializes in the subsequent period (thus the discounting) and works through the tax induced reduction in savings in all regions (note that $\mathrm{d}\ln(s_t)/\mathrm{d}\tau_t = -1/(1 - \tau_t - \tau_t^j)$, see Eq. (4)). The benefit arises in the form of higher interest rates (reflected in ϵ_{Rk}, which is negative), and the cost in the form of a lower tax base to fund public services in the future (reflected in ϵ_{wk}, which is positive) weighted by the preference for public services in the subsequent period, γ_{t+1}. We have the following result:

Proposition 1. *Suppose that $\epsilon_{Rk} + \epsilon_{wk}\gamma_{t+1} \neq 0$ such that $\mathcal{F}_t \neq 0$. Then, in equilibrium, only one level of government levies taxes. In particular, for $\epsilon_{Rk} + \epsilon_{wk}\gamma_{t+1} < 0$ (such that $\mathcal{F}_t > 0$) only the federal government levies taxes and for $\epsilon_{Rk} + \epsilon_{wk}\gamma_{t+1} > 0$ (such that $\mathcal{F}_t < 0$) only the regional governments levy taxes.*

The proof is immediate; we relegate it to the online appendix. Intuitively, the degree of tax centralization is determinate because voters at nationwide and regional elections perceive different net benefits of taxation. When lower savings drive up interest rates sufficiently strongly to render $\mathcal{F}_t > 0$, then the federal government levies taxes because voters at nationwide elections internalize that taxation improves their inter temporal terms of trade. In contrast, when lower savings depress next period's wages sufficiently strongly and the preference for public services in the subsequent period is sufficiently high to render $\mathcal{F}_t < 0$, then regional governments levy taxes because only voters at nationwide elections internalize the cost of taxation that results from lowering next period's tax base. A binding commitment for regions not to raise taxes would improve voters' welfare in the latter case.

We have determined the level of government that collects taxes, the equilibrium policies and thus, conditional on z_0, the competitive equilibrium. Note that conditions (9) and (10) and Proposition 1 verify the conjecture: The fact that the capital stock does not enter the first-order (and complementary slackness) conditions implies that the policy choices implied by conditions (9) and (10) are orthogonal to the endogenous state variables. That is, under the conjecture that the policy choices of future governments are orthogonal to the endogenous state, the implied policy choices of the contemporaneous governments are orthogonal to the endogenous state as well.[28]

Notwithstanding this orthogonality, the trade-offs underlying the conditions are dynamic in nature as they relate contemporaneous tax revenue and spending to future factor prices and revenue. The tractability of the model thus does not arise from suppressing this dynamic interaction, as in static models, but from specifying functional forms that render the factor price elasticities and the derivatives of the indirect utility functions orthogonal to the capital stock.[29]

The "bang-bang" property of the equilibrium policy is a direct consequence of the assumption that expenditure at the regional and federal level are perfect substitutes; the fact that only $\mathcal{F}_t$ drives a wedge between the regional and federal first-order conditions; and the fact that the sign of this wedge does not vary with taxes. In the next section, we introduce preference heterogeneity and spending complementarities. This smoothes the response of (average) regional taxes to an increase in $\mathcal{F}_t$.

To highlight the importance of general equilibrium effects for Proposition 1, we consider the implications of changing the tax base from labor to capital income. At the time when capital income taxes are decided upon and implemented, they only affect consumption of the old, but not savings of the young. As a consequence, the federal government perceives no general equilibrium factor price effects, $\mathcal{F}_t = 0$. Moreover, the weight that political candidates attach to the cost of taxation changes from $(1 + \beta)$ to ω/ν_t since the old rather than the young bear the tax burden. Otherwise, the first-order conditions for taxes remain unchanged. Total taxes thus are determinate in equilibrium but the degree of centralization of tax collections is not. For in an economy with capital income taxes, voters at the federal and regional level perceive exactly the same trade-off when weighing the pros and cons of a tax hike, unlike in the baseline model with labor income taxes.

The key message of Proposition 1 is that in politico-economic equilibrium, the degree of tax centralization depends on dynamic general equilibrium effects. This result is unchanged if we allow for productivity change. Due to our assumption of logarithmic utility and Cobb-Douglas technology, both the structure of the political first-order conditions and the factor price elasticities entering them are unaltered if productivity varies over time. In the online appendix, we establish that the finding also is robust along many other dimensions. We allow for endogenous labor supply, tax distortions, and additional

[28] The equilibrium is the unique symmetric Markov perfect equilibrium arising in the limit of the finite-horizon economy. See the discussion after definition 2 and the proof of Proposition 1 in the appendix.

[29] As shown elsewhere, these functional form restrictions tend to be of minor importance for the quantitive predictions of the model. While small deviations from logarithmic utility imply that equilibrium policies are non-trivial functions of the capital stock they nevertheless tend to generate very similar numerical predictions for the equilibrium outcomes (for an analysis in a related context, see Gonzalez-Eiras and Niepelt, 2005).

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

policy instruments and show that the result is robust since the perceived cost differences due to general equilibrium effects are unrelated to the effects of tax distortions. A similar result follows if we introduce labor mobility across regions.[30] The results also are robust when we introduce policy instruments for intergenerational redistribution, such as public debt or social security,[31] or if we assume that households are longer-lived.

5. Complementarities, heterogeneity, and grants

To generate a role for federal grants, we introduce complementarities between regional and federal spending. Accordingly, both the federal and the regional governments must spend resources for public services to be provided, and efficiency calls for the two levels of government to spend resources in specific proportions. By channeling revenue to the regions, federal grants decouple the composition of government spending across levels of government from the composition of government revenue collection. This decoupling is useful when the net benefit of taxation, $\mathcal{F}_t$, renders it "cheaper" for the federal government to tax.

We also introduce two key elements of fiscal federalism models, namely cross-regional static externalities from public service provision, and preference heterogeneity regarding public services. The presence of static externalities implies that the federal government might find it advantageous to subsidize regional government spending, thus introducing a second motive for federal grants. Heterogeneity, in turn, determines the strength of the motives for federal taxation and grants.

Assume, then, that there are two types of regions, with high and low preference for public services, γ_t^1 and $\gamma_t^2 < \gamma_t^1$ respectively. The share of high and low preference regions is denoted θ_t^1 and θ_t^2. In symmetric equilibrium all regions within the same group behave identically, except possibly a set of regions of measure zero. The endogenous state therefore contains the savings of retirees in the typical regions, $\vec{s}_{t-1} = (s_{t-1}^1, s_{t-1}^2)$, and equilibrium tax policy at date t is given by $(\tau_t^1, \tau_t^2, \tau_t)$.

The federal government pays a positive, uniform grant, x_t, to regional governments.[32] We allow for proportional deadweight losses of grants at rate $1-\sigma \geq 0$.[33] Accordingly, condition (3) generalizes to

$$e_t = w_t(\tau_t - x_t), \quad e_t^i = w_t(\tau_t^i + \sigma x_t), \quad e_t^j = w_t(\tau_t^j + \sigma x_t), \quad j = 1, 2. \tag{11}$$

To capture complementarities between regional and federal spending, we assume that public services in region i are given by

$$\begin{aligned} g_t^i &= \left[(e_t^i)^\delta (e_t)^{1-\delta}\right] \cdot \left[\left((e_t^1)^\delta (e_t)^{1-\delta}\right)^{\theta_t^1} \cdot \left((e_t^2)^\delta (e_t)^{1-\delta}\right)^{\theta_t^2}\right]^\lambda \\ &= w_t^{1+\lambda} (\tau_t^i + \sigma x_t)^\delta (\tau_t - x_t)^{(1-\delta)(1+\lambda)} \prod_{j=1}^{2} (\tau_t^j + \sigma x_t)^{\delta\lambda\theta_t^j}. \end{aligned}$$

Here, λ measures the strength of static, cross-regional externalities, and $\delta \in (0, 1)$. In the working paper version, we discuss micro foundations for this specification (Gonzalez-Eiras and Niepelt, 2017).[34] We also provide derivations and discussions for the general case with many types of regions.

Households' savings choices are unaffected by these changes and the definition of equilibrium is modified in the obvious way. The political first-order conditions with respect to τ_t^j and τ_t, respectively, now read

$$\left(\frac{\omega}{\nu_t} + 1\right) \frac{\gamma_t^j \delta}{\tau_t^j + \sigma x_t} - \frac{1+\beta}{1-\tau_t-\tau_t^j} \leq 0, \quad j = 1, 2, \tag{12}$$

$$\sum_{j=1}^{2} \theta_t^j \left\{ \left(\frac{\omega}{\nu_t} + 1\right) \frac{\gamma_t^j (1-\delta)}{\tau_t - x_t} - \frac{1+\beta}{1-\tau_t-\tau_t^j} \right\} + \mathcal{E}_t + \mathcal{F}_t \leq 0, \tag{13}$$

where $\mathcal{E}_t \equiv \lambda \bar{\gamma}_t (\omega/\nu_t + 1)(1-\delta)/(\tau_t - x_t)$ captures cross-regional externalities and "bars" (as in $\bar{\gamma}_t$) denote nationwide averages. Condition (12) differs from condition (9) in the baseline model because the marginal benefit of taxation only depends on regional spending, reflecting the assumption of complements rather than perfect substitutes (as well as logarithmic utility). Condition (13) differs from condition (10) for the same reason. Moreover, the summation over types in condition (13) reflects preference heterogeneity, and the cross-regional externalities ($\lambda > 0$) introduce the new $\mathcal{E}_t$ term and slightly modify the general equilibrium factor price effects, $\mathcal{F}_t$, which are now given by $-\beta(\epsilon_{Rk} + \epsilon_{wk}(1+\lambda)\bar{\gamma}_{t+1})/(1-\tau_t-\bar{\tau}_t)$.

[30] Tax distortions and labor mobility introduce additional, *static* effects on the degree of tax centralization.

[31] To establish this, we rely on results in Gonzalez-Eiras and Niepelt (2015).

[32] In the working paper, we analyze matching grants (Gonzalez-Eiras and Niepelt, 2017), concluding that the model predictions are qualitatively unaffected by the type of grant.

[33] In addition to capturing resource costs of inter governmental transfers, the "deadweight losses" might serve as stand in for other frictions or model specification errors.

[34] We discuss constitutional restrictions that prescribe which services must be provided (but not necessarily financed) by regional or federal governments. The division could reflect externalities, spillovers, or the strength of tax-benefit linkages for local voters, as highlighted by Tiebout (1956). See also Hatfield and Padró i Miquel (2012).

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

In addition to (12), (13), and the complementary slackness conditions a first-order condition for grants holds in equilibrium:

$$\sigma\delta\sum_{j=1}^{2}\frac{\theta_t^j(\gamma_t^j+\lambda\bar{\gamma}_t)}{\tau_t^j+\sigma x_t}-\frac{(1-\delta)(1+\lambda)\bar{\gamma}_t}{\tau_t-x_t}\leq 0. \tag{14}$$

The first term reflects the benefit from higher regional spending and the second term represents the cost due to lower federal spending. Note that conditional on taxes, the degree of preference *heterogeneity* but not the *average* preference for public services affects the choice of grants.[35] In fact, condition (14) implies that to be consistent with the joint evolution of taxes and grants in U.S. postwar data, the model necessarily requires time variation in preference heterogeneity.[36] We return to this point in Section 6.

The following proposition characterizes the politico-economic equilibrium in the extended model:

Proposition 2. *Consider the model with complementarities, heterogeneity, and grants. Define* $\Omega_t\equiv(\omega/\nu_t+1)$; $\Phi_t(\gamma_t)\equiv 1+\beta+\delta\Omega_t\gamma_t$; *and* $\Lambda_t\equiv(\epsilon_{Rk}+\epsilon_{wk}(1+\lambda)\bar{\gamma}_{t+1})=-\mathcal{F}_t(1-\tau_t-\bar{\tau}_t)/\beta$. *In equilibrium,*

(i) the federal government always levies taxes;

(ii) if

$$\Phi_t(\bar{\gamma}_t)+\frac{\frac{\beta}{1+\beta}\Lambda_t}{\sum_{j=1}^{2}\frac{\theta_t^j}{\Phi_t(\gamma_t^j)}}\geq\sigma\left(\Phi_t(\bar{\gamma}_t)+\lambda\bar{\gamma}_t\sum_{j=1}^{2}\frac{\theta_t^j\Phi_t(\gamma_t^j)}{\gamma_t^j}\right), \tag{15}$$

then all regions levy taxes as well and grants generically equal zero;

(iii) if the opposite condition holds, then grants are strictly positive and fully crowd out taxes in regions with a low valuation of public services;

(iv) more preference heterogeneity (for given $\bar{\gamma}_t$*) reduces the set of values for the other parameters for which condition (15) holds, rendering grants more likely.*

The proof is relegated to the online appendix. Recall that grants have value when it is "cheaper" to raise revenue at the federal level (because of the dynamic externalities) or when it is beneficial to subsidize regional spending (because of static, cross-regional externalities). Condition (15) reflects this. With homogeneous preferences ($\gamma_t^j=\bar{\gamma}_t$) and no deadweight losses ($\sigma=1$), the condition reduces to

$$1+\frac{\beta}{1+\beta}\Lambda_t\geq 1+\lambda,$$

highlighting the role of the two types of externalities: Positive static, cross-regional externalities ($\lambda>0$ and $\mathcal{F}_t=\Lambda_t=0$) or positive general equilibrium effects of taxation ($\lambda=0$ and $\mathcal{F}_t>0$ and thus, $\Lambda_t<0$) render it less likely that condition (15) is met and thus, more likely that grants are used. Preference heterogeneity and deadweight losses enrich this basic trade-off. With heterogeneity, grants crowd out taxation in regions with a low valuation of public services.[37] Deadweight losses reduce the net benefit of grants, as does less dispersion of preferences.

6. Quantitative analysis

We quantitatively assess the extended model's explanatory power based on predictions for the relative size of the federal government as well as for federal grants. We also assess the importance of dynamic general equilibrium effects for policy outcomes by computing counterfactuals and evaluate the plausibility of alternative explanations for the fiscal transformation in the U.S.

Focusing first on the evolution of revenue and grant shares, recall from Fig. 1 in the introduction that the share of the federal government dramatically rose between 1920 and 1950, with most of the increase occurring during the 1930s. During

[35] This follows from dividing (14) by $\bar{\gamma}_t$. A change in the average preference for public services affects grants indirectly, through its effect on tax rates, as reflected in conditions (12) and (13). But this is a second-order effect.

[36] To see this, suppose that grants are strictly positive (condition (14) holds as an equality) and regional taxes are strictly positive in one region, region 1 (see Proposition 2 below). Let $Z_t^j\equiv\theta_t^j(\frac{\gamma_t^j}{\bar{\gamma}_t}+\lambda)$, $j=1,2$, and $Z\equiv(1-\delta)(1+\lambda)/(\sigma\delta)$. Totally differentiating condition (14) (holding the Z_t^js constant) yields

$$dx_t\left[\frac{Z}{(\tau_t-x_t)^2}+\frac{Z_t^1\sigma}{(\tau_t^1+\sigma x_t)^2}+\frac{Z_t^2}{\sigma x_t^2}\right]=-d\tau_t^1\frac{Z_t^1}{(\tau_t^1+\sigma x_t)^2}+d\tau_t\frac{Z}{(\tau_t-x_t)^2},$$

which implies $-\sigma^{-1}<dx_t/d\tau_t^1<0$ and $0<dx_t/d\tau_t<1$. Note that this result also holds if preferences change over time as long as γ_t^j is proportional to $\bar{\gamma}_t$ and the regional shares are constant (because the Z_t^j terms are time invariant in this case). In the data regional tax rates almost double in the post-war period while the federal tax rate increases by less than 20%. Subject to constant preference heterogeneity the model therefore predicts—for otherwise arbitrary model parameters—at most a 20% increase in grants, which is counterfactual (see also Section 6).

[37] In the working paper, we analyze matching grants (Gonzalez-Eiras and Niepelt, 2017). We find that all regional tax rates are positive, with regions with low valuation for public services imposing lower tax rates.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

Table 1
Baseline calibration.

Parameter	Description	Value
α	Capital share in production	0.2815
β	Discount factor	0.6133
γ^1_{2000}	High valuation, 2000	0.8032
γ^2_{2000}	Low valuation, 2000	0.0223
$\gamma^j_{t+1}/\gamma^j_t$	Common growth rate of γ^j_t	1.1779
δ	Regional share in public service	0.4830
λ	Strength of static externality	0.0000
$1-\sigma$	Deadweight losses	0.0750
ω	Political weight of retirees	0.9176

See the text for explanations and sources.

the same period, federal grants started to gain importance as a source of state funding and since then, the share of federal grants in state and local revenues has continued to grow.[38]

In principle, the observed rise in the federal share of government and federal grants could be explained by time variation in deadweight losses (σ); static externalities (λ); the importance of federal vs. regional spending (δ); or preferences (γ^j_t and θ^j_t), see Proposition 2. Since there is little tangible evidence for variation of the former three factors we focus on the fourth factor, preferences, as a potential driver. Recall that the model predicts preference *heterogeneity* to affect the equilibrium choice of grants. Changes in the *average* preference for public services, which we allow for, cannot explain a significant trend increase in grants.

We associate regions with a strong or weak preference for public services with urban or rural regions, respectively, and we proxy the share of urban regions, θ^1_t, by the U.S. urbanization rate as reported by the Census Bureau. In the online appendix, we provide several pieces of evidence that support this association.

To calibrate the parameter values, we use five moment conditions as well as a-priori information. The first moment condition represents the Euler equation in steady state, relating β to the 30-year gross interest rate, $R = 2.443$ (see Gonzalez-Eiras and Niepelt, 2008). In the baseline calibration, we postulate no static externalities ($\lambda = 0$) and 7.5% deadweight losses ($\sigma = 0.9250$). We posit a Cobb-Douglas production function for the final good and set the capital share in the production function to $\alpha = 0.2815$, based on findings in Piketty and Saez (2003). From Gonzalez-Eiras and Niepelt (2008) we take $\omega = 0.9176$. We assume that one period in the model corresponds to 30 years in the data and associate ν_t and θ^1_t with the 30-year gross U.S. population growth rate and the share of urban regions as reported by the Census Bureau.

The remaining four moment conditions are based on the model's political first-order conditions. Recall that our theory explains the federal tax share as well as federal grants conditional on the evolution of total government revenue. Accordingly, the calibration imposes the long-term development of total government revenue but it does *not* constrain the evolution of revenue shares and grants. That is, all model predicted dynamics of grants and revenue shares are not pinned down by the moment conditions.

Specifically, we use the political first order conditions for federal taxes, regional taxes, and grants evaluated in the year 2000 to match the size of the federal government and of total government (defined as federal, state, and local spending relative to GDP) as well as the GDP-share of grants in that year. Moreover, we use the combined political first-order conditions for federal and regional taxes evaluated in the year 1950 to match the size of total government.[39,40] We assume that only high valuation regions impose regional taxes and thus, that only in high valuation regions the first-order condition with respect to the regional tax rate is interior. This yields four moment conditions. It also yields an inequality condition (reflecting the zero tax rate in regions with low valuation). We check that this inequality constraint is satisfied under our calibration, thus making sure that the assumptions underlying the calibration are internally consistent.

Table 1 summarizes the calibration and Fig. 3 illustrates the demographic trends that we feed into the model. The calibrated β value corresponds to an annual discount factor of approximately 0.9838. The calibration for δ suggests an almost equal importance of federal and regional spending in the provision of public services. In order for rural areas not to levy taxes the calibration assigns a very low value to the preference for public services in these regions, at approximately 3% of the value in urban areas.[41] Accordingly, the model predicts a counterfactually high ratio (7.0) of government spending in

[38] Between 1900 and 1930, the federal government's share in tax collections averaged 38.3%; by 1950 this share had risen to 72.6%. As a share of GDP, federal grants surpassed 0.5% in the 1950s and approached 2.8% in 2014.

[39] Data comes from the National Income and Product Accounts (NIPA) from the Bureau of Economic Analysis. In the model there is no public debt; revenues and expenditure thus are equivalent measures of the size of government. To account for the absence of debt in the model we use the average of current revenues and current expenditures as our measure of the size of government. For state and local governments we subtract federal grants. We include social security in our measure because we do not have sufficient information to correct for all other age specific expenditure components such as for education or health. The grant measure includes all federal grants-in-aid (including health, income security, education, training, transportation, community and regional development, etc.) received by state and local governments.

[40] We use 1950 rather than 1940 as the base year because by that time government finances likely reflected a more regular, post-recovery and post-war mode.

[41] This feature is robust to assuming matching grants, see Gonzalez-Eiras and Niepelt (2017).

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

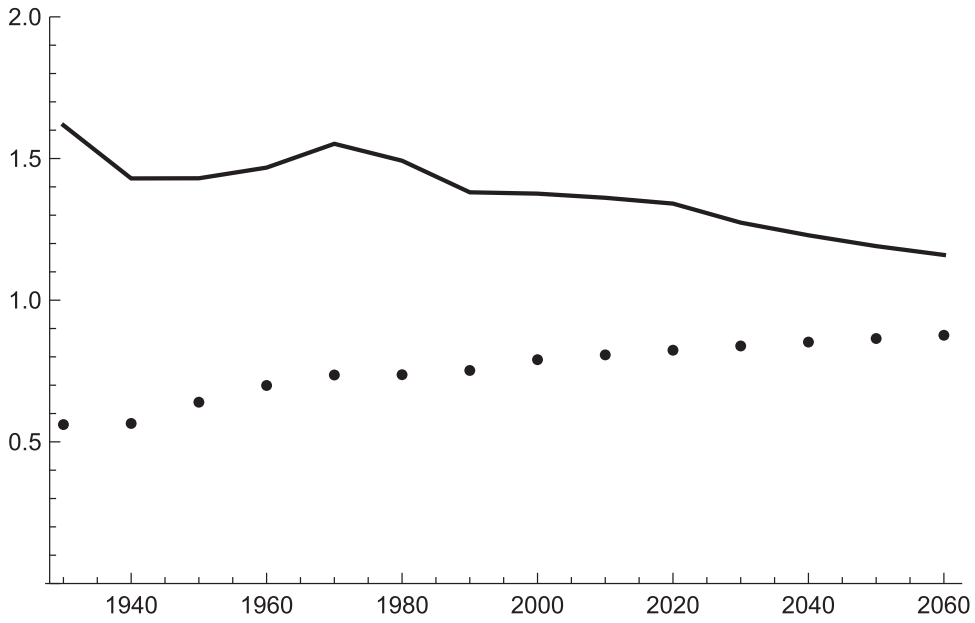

Fig. 3. Demographics and urbanization. Population growth rate over thirty years (solid) and urbanization (dots). Data from U.S. Census Bureau. Projections for population growth as reported by Census Bureau (middle series). Projections for urbanization interpolated based on United Nations (2014) forecast for 2050.

urban relative to rural areas. Introducing an exogenous component for regional tax collections easily resolves this problem without interfering with the other model predictions.[42] Finally, to replicate the increasing size of government between 1950 and 2000, the calibration requires the preference for public services to grow at about 0.55% per year or 17.8% over thirty years. This is qualitatively consistent with Wagner's law and with the evolution over time of attitudes towards spending cuts, as reported in the online appendix.[43]

By construction, the calibrated model perfectly predicts the size of total government in the years 1950 and 2000 as well as the federal share in tax collections in the year 2000. The predicted change of the federal share between World War II and the year 2000, which is not constrained by the calibration, is nearly zero, compared to a slight decline in the data.[44] As for federal grants, the model captures the long-term increase but not the short-run fluctuations (notably during the 1970s and the Great Recession), see Fig. 4. This reflects the fact that in the data, grants also are used for redistributive and risk sharing purposes which our model does not speak to. Importantly, the increase in grants since the 1930s reflects rising urbanization and thus, preference heterogeneity, not the stronger *average* preference for public services that the calibration imposes. This is particularly evident when we simulate the model subject to constant rural and urban shares (at their year 2000 values); the model then predicts a slightly negative trend for grants, see Fig. 4. In contrast to the important role played by urbanization, the changing demographics only are of minor importance. If we fix the population growth rate at its year 2000 value, the model predictions barely change.[45] Out of sample, the model predicts that grants continue to increase in the future up to approximately 4.9% of GDP in the year 2060.

To assess the robustness of the quantitative model predictions we check how different assumptions about the values of the parameters λ, σ, and ω affect the results. To this effect, we first re-calibrate the remaining model parameters based on the moment conditions described earlier. Thereafter, we use the newly calibrated model to generate predictions for grants and taxes both in and out of sample.

The model predictions are sensitive with respect to changes in the parameter values of λ or σ but much less so with respect to changes in the value of ω. When static externalities are negative, $\lambda = -3\%$ say, or deadweight losses higher, $\sigma = 90\%$ say, then the predicted grants peak at between 2.8 and 2.9% of GDP between 2040 and 2050 before reverting back

[42] In state level data from the Census Bureau, the ratio of government spending in urban relative to rural areas extends up to roughly 2.5. If we assume that regions exogenously collect a tax of 5% to fund spending unrelated to the provision of public services, then the predicted ratio of spending in urban relative to rural areas equals 2.5. This modification does not change the prediction of a trend increase in grants; in the modified model grants peak at 5.3% of GDP, instead of 4.9%, at the end of the simulation horizon.

[43] An alternative explanation for the rising size of government could rely on public services being a luxury good, and higher incomes. Our assumption of logarithmic preferences rules out income effects on tax rates.

[44] Baicker et al. (2012) document the post World War II increase in the share of state governments. They argue that the increase reflects changed incentives provided by federal policies.

[45] This reflects the fact that population growth does not directly enter the equilibrium condition for grants, Eq. (14), but only indirectly through its effect on taxation.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

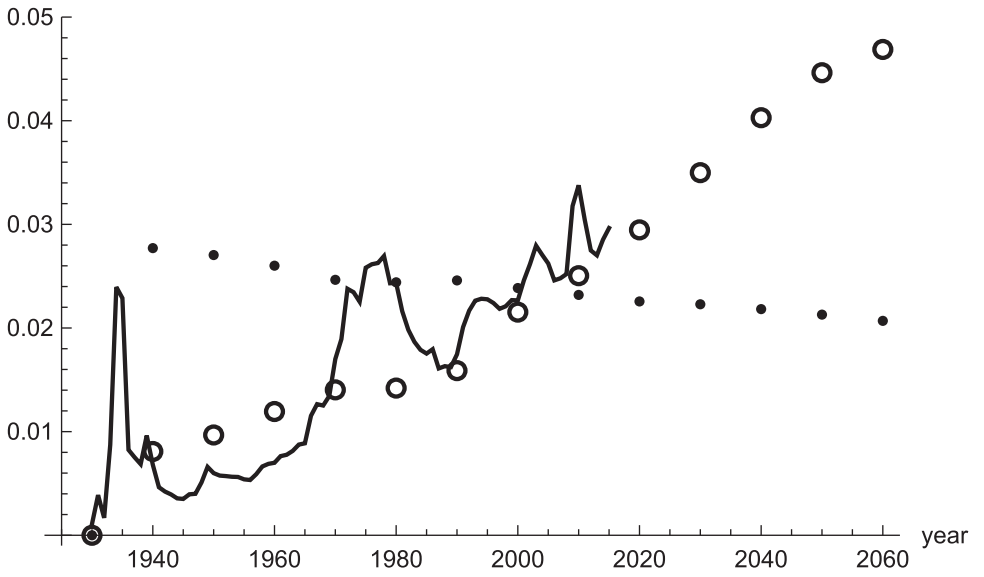

Fig. 4. Federal grants, share of GDP. Data from NIPA (solid), model predictions (circles), model predictions with constant urbanization (dots).

to lower values. When spending externalities are positive, $\lambda = 2\%$ say, or deadweight losses lower, $\sigma = 94\%$ say, then grants are predicted to increase to between 9.3 and 10% of GDP in 2060. Intuitively, with lower deadweight losses or higher static externalities, the federal government has a stronger incentive to provide grants (see Proposition 2).

In contrast, the parameter value of ω must be changed by much more (reflecting the more symmetric effect of ω in condition (15)) to generate the same predictions for grants. It would have to equal 0.6423 (1.1470) in order to generate grants that peak at 2.8% of GDP in 2050 (reach 9.2% of GDP by 2060). The calibration is not much affected by these changes in λ, σ, or ω, except for the preference parameter γ^2_{2000} (and also the parameter γ^1_{2000} in the case of ω).

To assess the importance of dynamic general equilibrium effects for the rise in federal taxation in the 1930s we proceed as follows: We use the calibrated model with dynamic general equilibrium effects from taxation ($\mathcal{F}_t \neq 0$) to predict the share of the federal government in the year 1950. We then shut down these general equilibrium effects ($\mathcal{F}_t = 0$) and solve for the politico-economic equilibrium; we associate that equilibrium with the situation before the fiscal transformation when the federal government mostly relied on tariffs and property taxes. We find that shutting down the dynamic general equilibrium channel implies a drop in the relative share of the federal government from 63.6% to 54.4%.[46] Since in the data the share of the federal government rises from 38.3% in the year 1930 to 72.6% in the year 1950, the model explains nearly 30% of the actual increase.

Under the assumption that federal and regional spending are perfect substitutes rather than complements, the predictive power of the model increases further: the model explains nearly all of the observed increase in the federal government's share. In the online appendix, we rationalize this finding in more detail. Since in our view, the model with spending complementarities yields a more realistic description of the provision of public services, we do not further pursue the quantitative analysis of the model in which federal and regional spending are perfect substitutes.

Finally, we assess the plausibility of alternative explanations for the rise in the federal government's revenue share. One potential explanation relates to changes in the structure of labor markets. Cole and Ohanian (2004) document how policies during the 1930s and 1940s first fostered and then reversed cartelization in labor markets and elsewhere (see, e.g., their section III.C), resulting in relatively strong wage growth in the 1930s and a dampened recovery. Labor cartelization would contemporaneously affect wages in our setup. but the wage effect would not asymmetrically enter the political first-order conditions and thus, could not explain the rise in the federal government's revenue share. In contrast, future cartelization could alter the anticipated factor price elasticities and thus, the general equilibrium effects that enter into the federal but not the regional first-order condition. But since the cartelization measures of the 1930s already were reversed a few years after their introduction, it appears unlikely that they could have significantly altered long-term expectations.[47]

More generally, explanations based on temporary economic shocks all suffer from the problem that they cannot easily explain why a temporary shock should have permanent effects on the structure of fiscal federalism. For example, a drop in property prices could have significantly affected the regional funding base and thus, put pressure on higher level governments to transfer resources. But since prices recovered and property returns typically are high (see, for example, Jordà et al.,

[46] The results are similar when we calculate the counterfactual imposing the demographic structure of the year 1930 rather than 1950.

[47] There is no evidence for a persistent change in the factor shares of labor and capital since 1929 (Piketty and Saez, 2003, Figure 6). In the model, altered factor price elasticities would require modified factor shares.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

2017) it is not clear why a property price shock should have permanently changed the U.S. fiscal system. Nor would such a shock explain why transfers increased; alternatively, the federal government or state governments could have assumed spending responsibilities from states and municipalities, respectively.

Another potential explanation relies on a stipulated increase in static externalities from regional spending.[48] To evaluate the ability of this alternative mechanism to quantitatively explain the fiscal transition in the U.S., we re-calibrate the model under the assumption that the federal government does not perceive the dynamic general equilibrium effects of labor income taxation (or that they are not present), $\mathcal{F}_t = 0$, and that the fiscal transformation is the result of a permanent change in λ in the 1930s. We find that the required change in static externalities is implausibly large, or that the increase of federal grants predicted by this alternative model fits the data less well.[49] We conclude that dynamic general equilibrium effects of taxation offer an explanation for the U.S. fiscal transformation whose quantitative implications compare well with an explanation based on static externalities.

7. Concluding remarks

What determines the degree of centralization of tax collections in a federal union? We propose a novel explanation that stresses differences in the perceived cost of taxation across levels of government due to dynamic general equilibrium effects. The dynamic externalities we emphasize complement static externalities that have traditionally been analyzed in the fiscal federalism literature, including spending externalities and externalities from horizontal or vertical tax competition.

When augmented with government spending complementarities our model also generates a role for inter governmental grants. Grants have value when they allow to channel revenue from the federal government where tax revenue is “cheap” to regions, or when regions underspend because they do not internalize positive cross-regional externalities.

We find that dynamic general equilibrium effects can help explain the U.S. fiscal transformation during the 1930s towards more centralized revenue collection, more widespread use of grants, and increased reliance on income taxation. In our framework these changes result in response to the ratification of the Sixteenth Amendment, which opened the door for federal labor income taxation in response to higher demand for government expenditure, specifically New Deal policies and World War II spending.

Our simple framework abstracts from cross-regional insurance, redistribution, and many other features that are present in federalist states. Given this simplicity, the model’s quantitative performance is reassuring. The model with spending complementarities accounts both for the trend increase in federal grants since 1930 and for roughly 30% of the observed increase in the relative size of the federal government during the 1930s.

Two extensions of the model appear to be of particular interest. First, the setup could be extended to admit productivity differences across regions, generating a role for cross-regional insurance and redistribution. Such an extension could be used to study the determinants of redistributive federal grants and the consequences of cross-regional inequality, for instance in the post-World-War II U.S. or in the context of European integration.

Second, the option to issue government debt for tax smoothing or tax burden shifting purposes could be introduced. Governments would hold conflicting views about the costs and benefits of public debt since regional policymakers would not internalize the general equilibrium effects of deficits on interest rates and wages. As a consequence, the federal government might opt to employ grants (and deficits) to influence both regional taxes and deficits. This extension could address questions about debt and deficit policies in federal states.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmoneco.2019.02.007.

References

Agranov, M., Palfrey, T.R., 2015. Equilibrium tax rates and income redistribution: a laboratory study. J. Public Econ. 130, 45–58.
Alesina, A., Spolaore, E., 1997. On the number and size of nations. Q. J. Econ. 112 (4), 1027–1056.
Baicker, K., Clemens, J., Singhal, M., 2012. The rise of the states: U.S. fiscal decentralization in the postwar period. J. Public Econ. 96 (11–12), 1079–1091.
Besley, T., Persson, T., 2009. The origins of state capacity: property rights, taxation, and politics. Am. Econ. Rev. 99 (4), 1218–1244.
Boffa, F., Piolatto, A., Ponzetto, G.A.M., 2016. Political centralization and government accountability. Q. J. Econ. 131 (1), 381–422.
Cole, H.L., Ohanian, L.E., 2004. New deal policies and the persistence of the Great Depression: a general equilibrium analysis. J. Polit. Econ. 112 (4), 779–816.
Dal Bó, E., Dal Bó, P., Eyster, E., 2018. The demand for bad policy when voters underappreciate equilibrium effects. Rev. Econ. Stud. 85 (2), 964–998.

[48] Maybe the most plausible candidate as a source of increased static externalities is public infrastructure investment to support major technological innovations. But many of these innovations (in particular, electric light and the internal combustion engine) already occurred at the end of the nineteenth century (Gordon, 2012). And by the time of the Great Depression, most of the infrastructure investments supporting them were already undertaken, at least in urban areas. The shifts in the fiscal landscape thus should have occurred earlier. Even if spending externalities had increased around the 1930s, federal spending should have spiked rather than permanently increased since the higher externalities would have triggered a federal public investment boom followed by more moderate maintenance spending. This is not what we see in the data.

[49] Subject to the δ value reported in Table 1, the calibrated λ value equals 0.3029 and the calibrated value for γ^2_{2000} is negative. When we drop the moment condition for grants in the year 2000 and impose for γ^2_{2000} the value reported in Table 1 then the calibrated value for λ equals 0.1246 and the predicted value for grants in the year 2000 exceeds the actual value by 33%.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007

Gonzalez-Eiras, M., Niepelt, D., 2005. Sustaining Social Security. Working Paper 1494. CESifo.
Gonzalez-Eiras, M., Niepelt, D., 2008. The future of social security. J. Monet. Econ. 55 (2), 197–218.
Gonzalez-Eiras, M., Niepelt, D., 2015. Politico-economic equivalence. Rev. Econ. Dyn. 18 (4), 843–862.
Gonzalez-Eiras, M., Niepelt, D., 2017. Fiscal Federalism, Grants, and the U.S. Fiscal Transformation in the 1930S. Discussion Paper 17–18. University of Copenhagen, Department of Economics.
Gordon, R.H., 1983. An optimal taxation approach to fiscal federalism. Q. J. Econ. 98 (4), 567–586.
Gordon, R.J., 2012. Is U.S. Economic Growth over? Faltering Innovation Confronts the Six Headwinds. Working Paper 18315. NBER, Cambridge, Massachusetts.
Greenstone, M., Looney, A., 2012. The Truth about Taxes: Just about Everyone Pays Them. The Hamilton Project. Brookings, Washington. http://goo.gl/dvnr0d.
Harstad, B., 2007. Harmonization and side payments in political cooperation. Am. Econ. Rev. 97 (3), 871–889.
Hatfield, J.W., Padró i Miquel, G., 2012. A political economy theory of partial decentralization. J. Eur. Econ. Assoc. 10 (3), 605–633.
Hilt, E., Rahn, W.M., 2018. Financial Asset Ownership and Political Partisanship: Liberty Bonds and Republican Electoral Success in the 1920s. Wellesley and University of Minnesota.
Jones, C.C., 1988. Class tax to mass tax: the role of propaganda in the expansion of the income tax during world war ii. Buffalo Law Rev. 37 (3), 685–738.
Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M., Taylor, A.M., 2017. The Rate of Return on Everything, 1870–2015. Working Paper 24112. NBER, Cambridge, Massachusetts.
Kessler, A.S., 2014. Communication in federal politics: universalism, policy uniformity, and the optimal allocation of fiscal authority. J. Polit. Econ. 122 (4), 766–805.
Kincaid, J. (Ed.), 2011. Federalism. Sage Publications, Los Angeles.
Knight, B., 2002. Endogenous federal grants and crowd-out of state government spending: theory and evidence from the federal highway aid program. Am. Econ. Rev. 92 (1), 71–92.
Kotlikoff, L.J., Rosenthal, R.W., 1990. Some Inefficiency Implications of Generational Politics and Exchange. Working Paper 3354. NBER, Cambridge, Massachusetts.
Krusell, P., Quadrini, V., Ríos-Rull, J.V., 1997. Politico-economic equilibrium and economic growth. J. Econ. Dyn. Control 21 (1), 243–272.
Lindbeck, A., Weibull, J.W., 1987. Balanced-budget redistribution as the outcome of political competition. Public Choice 52, 273–297.
Lockwood, B., 2002. Distributive politics and the costs of centralization. Rev. Econ. Stud. 69 (2), 313–337.
Mehrotra, A.K., 2013. Making the Modern American Fiscal State: Law, Politics, and the Rise of Progressive Taxation, 1877–1929. Cambridge University Press, New York.
Oates, W.E., 1972. Fiscal Federalism. Harcourt Brace Jovanovich, New York.
Oates, W.E., 1999. An essay on fiscal federalism. J. Econ. Lit. 37 (3), 1120–1149.
Persson, T., Tabellini, G., 1996. Federal fiscal constitutions: risk sharing and redistribution. J. Polit. Econ. 104 (5), 979–1009.
Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913–1998. Q. J. Econ. 118 (1), 1–39.
Rhode, P.W., Strumpf, K.S., 2003. Assessing the importance of Tiebout sorting: local heterogeneity from 1850 to 1990. Am. Econ. Rev. 93 (5), 1648–1677.
Soares, J., 2005. Public education reform: community or national funding of education? J. Monet. Econ. 52 (3), 669–697.
Song, Z., Storesletten, K., Zilibotti, F., 2012. Rotten parents and disciplined children: a politico-economic theory of public expenditure and debt. Econometrica 80 (6), 2785–2803.
Tiebout, C.M., 1956. A pure theory of local expenditures. J. Polit. Econ. 64 (5), 416–424.
Treisman, D., 2007. The Architecture of Government. Cambridge University Press, Cambridge.
United Nations, 2014. World urbanization prospects: the 2014 revision, highlights (ST/ESA/SER.a/352). Department of Economic and Social Affairs, Population Division, New York.
Vélez, J.L., 2014. War and Progressive Income Taxation in the 20th Century. Working Paper 2014-3. Berkeley Economic History Laboratory, Berkeley, California.
Wallis, J.J., 2000. American government finance in the long run: 1790 to 1990. J. Econ. Perspect. 14 (1), 61–82.
Wallis, J.J., Oates, W.E., 1998. The impact of the New Deal on American federalism. In: Bordo, M.D., Goldin, C., White, E.N. (Eds.), The Defining Moment: The Great Depression and the American Economy in the Twentieth Century. University of Chicago Press, Chicago, pp. 155–180.

Please cite this article as: M. Gonzalez-Eiras and D. Niepelt, Dynamic tax externalities and the U.S. fiscal transformation, Journal of Monetary Economics, https://doi.org/10.1016/j.jmoneco.2019.02.007