

A theory of top income taxation and social insurance*

Francisco M. Gonzalez and Jean-François Wen

October 2013

Abstract

The development of the welfare state in the Western economies between 1930 and 1990 coincided with a puzzling pattern in the taxation of top incomes. Effective tax rates at the top increased sharply but then gradually decreased, even as social transfers continued rising. We propose a new theory of the development of the welfare state to explain these facts. Our main insight is that social insurance and top income taxation are substitutes for averting social conflict. We emphasize the role of the Great Depression as a source of aggregate risk, and argue that the rise of the welfare state can be understood as a process of exploiting efficiency gains in response to gradual technological improvements in the provision of social insurance. Our detailed arguments build on the policy histories of the United States, Great Britain, and Sweden.

Keywords: supply of social insurance, taxation of top incomes, social conflict.

JEL classification: D30, H20, H50, P16.

*Gonzalez: University of Waterloo (francisco.gonzalez@uwaterloo.ca). Wen: University of Calgary (wen@ucalgary.ca). We have benefited from comments by Robin Boadway, John Burbidge, Herb Emery, Merwan Engineer, Wojciech Kopczuk, Ken McKenzie, John Roemer and Dan Usher. The first author gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.

1 Introduction

Social transfers grew spectacularly in the Western economies between 1930 and 1990. Meanwhile, tax rates on the top 1 percent of the income distribution increased sharply in the 1930s and 1940s but then decreased gradually between 1960 and 1990. This pattern of top income taxation is particularly striking in the United States and the United Kingdom, but it is qualitatively similar in other countries.¹ Why was there so little change in social transfers until the Great Depression? If the state taxes the wealthy to supply transfers to others, why did top income taxes decline after 1960?

We propose a new theory of the development of the welfare state to address these questions. A risk neutral elite, or *governing group*, is presumed to have control over policy, but it requires the cooperation of a risk averse *non-governing group* in order to be able to produce. The interests of the governing group are aligned with the economic interests of the top of the income distribution.² We characterize public policy as the outcome of Pareto efficient equilibria of a repeated game where in every period the governing group designs income transfers and the non-governing group makes a participation choice. The threat of conflict — non-participation in production — enables the non-governing group to obtain a part of the surplus from cooperation. Since the non-governing group is exposed to risks that private insurance markets fail to cover, the governing group has an incentive to provide redistribution in the form of state-contingent transfers, i.e., social insurance. This allows them to secure social peace with smaller average transfers compared to providing pure redistribution.

Our main insight is that insurance and top income taxation are substitutes for averting social conflict. This can explain the pattern of top income taxes and social transfers. Future incomes were perceived as riskier in the aftermath of the Great Depression, making social conflict more attractive. As social insurance coverage was initially inadequate for counteracting risk, an increase in redistribution financed by top income taxes was required to avoid conflict. We argue that constraints in revenue collection, record-keeping, and program ad-

¹See, e.g., Kuznets (1963), Piketty and Saez (2007) and the discussion in Atkinson (2004).

²For evidence that political influence rises sharply with income even in democratic societies, see Benabou (2000) and Gilens (2005). See Dal Bó et al. (2009) for evidence on the persistence of political elites in the United States. See Putnam (1976) for a discussion of political elites in democratic societies.

ministration necessitated, historically, an incremental approach to social insurance coverage. Consequently, the increase in top income taxation was reversed gradually as social insurance programs matured to provide better coverage, and conflict was made increasingly less attractive.

Through the lens of our theory, the expansion of the welfare state can be understood as a process of exploitation of efficiency gains made possible by learning-by-doing and technical improvements in public administration. This mechanism could not have operated before the taxation innovations resulting from World War II, nor after the 1980s, at which time all major social insurance programs were in place, and public administration records were fully computerized in the Western economies. Our view complements other explanations of the development of the welfare state, which are discussed in Section 1.

In the 1951 annual report of the National Bureau of Economic Research, Arthur Burns (1951: 3) described the unprecedented decline in the top income shares between 1930 and 1950 as “a transformation that has been carried out peacefully and gradually, but which may already be counted as one of the great social revolutions of history.”³ From our perspective, it is precisely the supply of social insurance by power elites who are *not* the main beneficiaries of the programs, and the increased taxation of the top incomes, that underlaid this *peaceful* transformation in the distribution of national income.

It is commonly argued that the increase in top income taxes in the period 1930–1950 was a key factor preventing the top income shares from recovering after World War II, and also that the tax cuts after the 1960s precipitated the increases in the income shares of the very top of the distribution in several countries after 1980.⁴ These arguments, however, take as given the evolution of taxation, without addressing its causes. Our view is that the period of declining top tax rates was the concomitant of a more efficient supply of social insurance.

In the rest of the paper, we develop the mapping between our formal theory and the history of welfare policy in the United States, Britain, and Sweden. In Section 2 we document the facts that motivate our theory, and we discuss the related literature. In Section 3 we

³Cited in Lindert (1998).

⁴See, e.g., Kuznets (1963), Piketty and Saez (2003), Saez (2004), and Atkinson and Leigh (2010).

present our basic model of social policy. In Section 4 we extend it to formalize the effects of changes in aggregate risk and in insurance opportunities. In Section 5 we discuss, for each country, the history of social insurance and tax legislation in relation to the threat of conflict and to the role of elites in the development of the welfare state. Section 6 concludes. Technical proofs are in the Appendix.

2 Stylized facts and related literature

The expansion of social insurance coverage in the Western economies is reflected in the growth of government transfers to individuals. Figure 1 depicts the remarkable growth of social transfers between 1930 and 1990. Social transfers as a share of GDP in 1930 were only 0.6, 2.2 and 2.6 percent in the US, Britain and Sweden, respectively. By 1990 these shares were 13.4, 19.5, and 30.8.⁵ Over this period, about three-quarters of transfers every year are in the form of social insurance payments.

[Figure 1]

Meanwhile, tax progressivity increased sharply in the 1930s and 1940s but then decreased gradually between 1960 and 1990. This pattern is reflected in the effective tax rates at the top of the income distribution, measured as the difference between the pre-tax income share and the post-tax income share, as a proportion of the pre-tax income share. Figure 2 depicts effective tax rates for the top 1 percent for the United States, Britain, and Sweden.⁶

[Figure 2]

Due to data limitations, Figure 2 plots effective tax rates for the top 1 percent. However, it should be noted that the evidence for the United States for the period 1960–2000 (Piketty

⁵Source: Lindert (2004, Table I.2) for 1900–1930; OECD Social Expenditure Database (1960–1981) and (1980–2000). A break in the series makes the post-1980 series not directly comparable to the pre-1980 series.

⁶We calculated the effective tax rates using data on pre-tax and post-tax income shares for the U.S., from Kuznets (1953) for the period 1930–1946 and from Piketty and Saez (2007) for the period 1960–2000; for the U.K., from Atkinson (2007) for the period 1937–2000; and for Sweden, from Statistics Sweden (2008) for the period 1975–2000. The data sources are not fully comparable. Kuznets (1953) refers only to income taxation. The post-tax income shares in Piketty and Saez (2007) are net of personal, corporate and estate taxes, and payroll taxes. The shares in the UK and Sweden are net of personal income taxes and social security contributions only.

and Saez, 2007) and Britain for the period 1937–2000 (Atkinson, 2007) indicates that the sharp increase in tax progressivity at the top after 1930 and the subsequent gradual decline that is apparent in Figure 2 is most evident for the upper fractiles of the top 1 percent.⁷ Data on post-tax income shares for the top 1 percent in Sweden is available to us only after 1975. However, United Nations data reported in Kuznets (1963) indicates that effective tax rates for the top 5 percent in Sweden increased from 8.9 to 15 percent between 1935 and 1948.

Conventional theories of the welfare state stress the demand for redistribution from rich to poor as a function of the political strength of the working class. The process of democratization is one such force that may lead to an increase in demands for redistribution. However, while Acemoglu and Robinson (2000) argue that earlier redistributive spending is likely to have been associated with the rise of democracy, the achievement of near universal suffrage before 1930 makes it difficult to understand how the rise of democracy by itself could be the main source of the rise in transfers after 1930 (Lindert, 2004: ch. 7, Aidt et al., 2006).

A related hypothesis is that the burden of the Great Depression and World War II fell disproportionately on the bottom half of the income distribution, and worsening inequality led subsequently to larger demands for redistribution, in the spirit of Meltzer and Richard (1981).⁸ However, the evidence does not support this hypothesis. In particular, the growth in social transfers coincided with *reduced* pre-tax income inequality.⁹

More generally, shifts in the political power of the working class could have affected tax progression and the size of transfers (Korpi, 1980, Korpi and Palme, 2003). However, this would not explain why the evolution of social transfers did not just mimic the pattern of top income taxes. It also fails to recognize that the power of the working class may itself be affected by aggregate risk. For instance, Rodrik (1998) provides evidence that countries

⁷The consensus is that the pattern of top income shares does not merely reflect patterns of tax avoidance and evasion, but rather changes in real incomes. See Atkinson (2004) and the references in footnote 4.

⁸The period of high tax progressivity cannot be attributed to the cost of World War II alone. For instance, top statutory tax rates remained high into the 1960s whereas, in contrast, after World War I the top tax rates fell in each country within five years after the war.

⁹For instance, the Gini coefficient in the United States was stable at 0.35–0.36 between 1955 and 1980 (U.S. Census Bureau, Historical Income Tables); in the United Kingdom it fell from 0.41 in 1949 to 0.37 in 1975 (Lowe, 1999); and in Sweden it fell from 0.38 in 1951 to 0.32 in 1973 (Bjorklund and Palme, 2000).

that are exposed to more risk spend more on social transfers.

The literature on the welfare state stresses the efficiency losses from redistributive policies. However, the slowdown of the welfare state in the 1990s is hard to understand solely in terms of the limits to social demands for redistribution. In particular, the view that the welfare state was simply unsustainable after 1980, because of the distortionary effects of taxes and transfers, is not well supported by the evidence (Atkinson, 1995, Lindert, 2004).¹⁰

Whereas the dominant approach for understanding the welfare state stresses redistribution as the essence of social policy, a second approach identifies the provision of insurance as primary. In our view, combining both approaches improves our understanding of the welfare state. It is apparent not only that many transfer programs were designed specifically to provide insurance, but also that they tend to be redistributive across individuals *ex ante* (Barr, 2004). Whether redistribution or insurance is viewed as central matters for assessing the consequences of the welfare state, which may be driven by efficiency losses from redistributive policies (Alesina and Rodrik, 1994, Persson and Tabellini, 1994), or by efficiency gains from protecting individuals against otherwise uninsurable risks (Benabou, 2000). In this paper, we argue that it was the gradual exploitation of efficiency gains from social insurance that drove the development of the modern welfare state, and that this process reflected the substitutability between top income taxation and insurance for averting social conflict.

3 A model of social policy

We formalize social policy as a self-enforcing contract in a repeated game between two groups, denoted A and B . Each group has a continuum of agents with identical mass normalized to 1. We assume that all agents within a group are identical, and they have resolved all free-riding problems so they act as a group. Group A is the *governing group*. By this we mean that it has the power to choose per capita transfers T^t to the *non-governing group* (group B) every period $t \geq 1$. We use the word *transfer* to denote generically any net income flows (positive or negative) from the governing group toward the non-governing group. From the

¹⁰See also Benabou (1996), and Romer and Romer, 2012)

perspective of the latter, a positive transfer is a subsidy and a negative transfer is a tax. It shall become clear that our arguments extend to the case of $N \geq 1$ non-governing groups.

The governing group's power is limited by the fact that the non-governing group chooses whether to participate in the social contract or to opt out. Opting out of the social contract is meant to formalize the consequences of social conflict in a simple manner. The important feature is that conflict is an inefficient outcome. If group B chooses not to participate in the social contract in period t , we assume that $y_A^t = y_B^t = 0$, where y_A^t and y_B^t denote the pre-transfer income of each individual in group A and group B , respectively, in period t .

Instead, if the non-governing group participates in the social contract, we assume that the pre-transfer per capita incomes of each group participating in the social contract are exogenous random variables, independent over time and across groups. Specifically, the pre-transfer income of each individual in group A in period t , $y_A^t > 0$, is drawn from some distribution $F_A(y_A^t)$ with positive support, whereas the pre-transfer income for each individual in group B is a weighted sum of $J > 1$ risky components $y_j^t > 0$, for $j = 1, \dots, J$. For simplicity, we assume that they have a common distribution $F(y_j^t)$, with positive support.

An important assumption is that the governing group can make the transfers T^t conditional on the realizations of y_A^t and on the realizations of a subset $K \subseteq \{1, \dots, J\}$ of the risky components y_j^t . Each factor $j \in K$ can thus be described as insurable, while each factor $j \notin K$ is uninsurable. Formally, the pre-transfer income y_B^t of each agent in group B is

$$y_B^t = \sum_{j \in K} a_j y_j^t + \sum_{j \notin K} a_j y_j^t, \quad (1)$$

where $a_j > 0$ are constant weights, and y_j^t is independent across risks and time periods.

We refer to the transfer schedule $T^t(y_A, \{y_j\}_{j \in K})$ as the *social contract* in period t , where T^t is a function of the realizations $\{y_A, \{y_j\}_{j \in K}\}$ of the random variables $\{y_A^t, \{y_j^t\}_{j \in K}\}$. For simplicity, we assume that agents in the non-governing group can never be made to pay an amount greater than their insurable income, that is,

$$T^t(y_A, \{y_j\}_{j \in K}) \geq - \sum_{j \in K} a_j y_j. \quad (2)$$

Thus, each factor $j \in K$ is both insurable and taxable, while each factor $j \notin K$ is both uninsurable and non-taxable. This may be interpreted as reflecting the fact that only the realizations of the insurable/taxable risk factors are verifiable.

Individuals in the non-governing group are risk averse, with per-period utility

$$u(y_B^t + T^t) = \frac{(y_B^t + T^t)^\theta}{\theta}, \quad (3)$$

with $\theta \in (0, 1)$, whenever they participate in the social contract. Since we have normalized conflict payoffs to zero, we restrict the coefficient of relative risk aversion $(1 - \theta)$ to be less than unity so there can be gains from trade. We also assume that agents in the governing group are risk neutral, which is a good approximation when the members of the governing group are sufficiently rich, or they have access to insurance markets, or they can self-insure.

3.1 The stage game

The timing of the stage game is as follows. Before the realization $\{y_A, \{y_j\}_{j \in K}, \{y_j\}_{j \notin K}\}$ of the shocks $\{y_A^t, \{y_j^t\}_{j \in K}, \{y_j^t\}_{j \notin K}\}$ is observed at date t , group A chooses a social contract $T^t(y_A, \{y_j\}_{j \in K})$, and simultaneously group B chooses whether to participate in the social contract or to opt out, where $x^t = 1$ if group B participates at date t and $x^t = 0$ if group B opts out. Then, if $x^t = 0$, the social contract breaks down, and all individuals receive a zero payoff. Instead, if $x^t = 1$, actual transfers take place according to $T^t(y_A, \{y_j\}_{j \in K})$, and individuals in groups A and B enjoy (*ex post*) payoffs $y_A - T^t$ and $u(y_B + T^t)$, respectively. For concreteness it may be useful to think of the governing group as offering a take-it-or-leave-it social contract every period, even though our analysis abstracts from the details of the institutional setting.

The set of feasible individually rational (*ex ante*) payoffs in the stage game is given by

$$V_A = \left[0, E \left(y_A^t + \sum_{j \in K} a_j y_j^t \right) \right], \quad (4)$$

and

$$V_B = \left[E \left(u \left(\sum_{j \notin K} a_j y_j^t \right) \right), E \left(u \left(y_A^t + \sum_{j \in K} a_j y_j^t + \sum_{j \notin K} a_j y_j^t \right) \right) \right], \quad (5)$$

for group A and group B , respectively, at any date t , where E is the expectation operator.

If the stage game is played just once, the unique Nash equilibrium, denoted by (x_{NE}, T_{NE}) , is such that the non-governing group participates in the social contract and the governing group appropriates all insurable/taxable incomes:

$$(x_{NE}, T_{NE}) = \left(1, - \sum_{j \in K} a_j y_j \right). \quad (6)$$

Group B does not opt out of the social contract, even though individuals anticipate that their insurable/taxable incomes will be expropriated, because their noninsurable/nontaxable income is positive whereas the conflict payoffs have been normalized to zero.

Our main results below would remain unchanged under alternative assumptions leading to a Nash equilibrium of the stage game where group B opts out of the social contract. Furthermore, our assumption that group A has all the bargaining power in the stage game serves to highlight the implications of social cooperation when the groups interact repeatedly, but it is otherwise inessential. What is important is that (1) group A can influence social transfers and thus the distribution of the social surplus in the insurable/taxable states and (2) the threat of inefficient conflict endows group B with some *de facto* bargaining power, the degree of which is endogenous to the equilibrium of the repeated game.

3.2 The repeated game

We shall focus on efficient perfect equilibria in pure strategies of the infinitely repeated game. To fix terminology, note that an *action profile* at date t , denoted by (x^t, T^t) , consists of a participation choice and a social contract. A *history* of the game at date t , denoted by $h^t = \{(x^\tau, T^\tau, Y^\tau)\}_{\tau=0}^{t-1}$, is the sequence of action profiles and realizations of the shocks $\{y_A^\tau, \{y_j^\tau\}_{j \in K}, \{y_j^\tau\}_{j \notin K}\}$, denoted by Y^τ , prior to date t . The game starts at date 0 with the null history h^0 . A *pure strategy profile* (σ_A, σ_B) is a sequence of functions $\sigma_i = \{\sigma_i^t(h^t)\}_{t \geq 0}$,

for $i = A, B$, mapping possible period- t histories to non-mixed actions.

A stream of action profiles, denoted by $Q = \{(x^t, T^t)\}_{t \geq 0}$, is referred to as a *path*, and Ω denotes the set of feasible paths. A strategy profile generates a path that can be defined inductively in the obvious manner. Preferences over paths for each individual in group A and B are determined according to the normalized utility functions

$$U_A(Q) = (1 - \delta) \sum_{t=0}^{\infty} \delta^t x^t E(y_A^t - T^t), \quad (7)$$

and

$$U_B(Q) = (1 - \delta) \sum_{t=0}^{\infty} \delta^t x^t E(u(y_B^t + T^t)), \quad (8)$$

respectively, where the discount factor $\delta < 1$ is common to all individuals. The normalization factor $(1 - \delta)$ implies that the normalized value of 1 util per period is 1, so payoffs in the stage game and in the repeated game are measured in the same units.

A pure strategy profile (σ_A, σ_B) is a (subgame) *perfect equilibrium* of the repeated game if, for every history h^t , the restriction of (σ_A, σ_B) to the histories consistent with h^t is a Nash equilibrium of the game from period t on with history h^t . A path Q is *Pareto efficient* if $U_i(Q) \geq U_i(Q')$, for $i = A, B$, for all paths $Q' \in \Omega$. A perfect equilibrium with the property that it implements a Pareto efficient path is referred to as an *efficient perfect equilibrium*.

3.3 Efficient perfect equilibria

In this section we characterize the properties of efficient perfect equilibria. We begin by characterizing the set of Pareto efficient paths, and then we show how they can be implemented by subgame perfect equilibria.

The set of Pareto efficient paths can be characterized by maximizing group A 's utility over all paths subject to group B 's utility being no lower than a given value v , for each feasible value v :

$$\max_{Q \in \Omega} \{U_A(Q)\} \text{ subject to } U_B(Q) \geq v; v \in V_B, \quad (9)$$

where V_B , $U_A(Q)$ and $U_B(Q)$ are given by equations (5), (7) and (8), respectively. The following result is easy to verify.

Proposition 1 *A path $Q = \{(x^t, T^t)\}_{t \geq 0}$ with $U_B(Q) = v$ is Pareto efficient if and only if it is a stationary path such that every period the non-governing group participates ($x^t = 1$ for all $t \geq 0$) and the social contract provides full insurance against insurable shocks by specifying the stationary transfer schedule $T^t = T_e$ for all $t \geq 0$, where*

$$T_e = c - \sum_{j \in K} a_j y_j,$$

and c is a constant satisfying

$$E \left(u \left(c + \sum_{j \notin K} a_j y_j^t \right) \right) = v \in V_B.$$

Proposition 1 implies that any feasible individually rational payoffs in the stage game can be generated by a Pareto efficient path. Note that receiving the payoff v every period amounts to a *normalized* sum of discounted utilities $U_B(Q)$ equal to v .

To gain some intuition, note that a solution to the problem given by (9), with $x^t = 1$ for all t , satisfies

$$1 = \lambda E \left(u' \left(y_B^t + T^t \right) | y_A, \{y_j\}_{j \in K} \right), \quad (10)$$

for all insurable/taxable income realizations $\{y_A, \{y_j\}_{j \in K}\}$, and for all $t \geq 0$, where λ is the Lagrange multiplier for group B 's participation constraint, and u' denotes marginal utility. That is, at any date $t \geq 0$, group A trades off, state by state, the marginal benefit from taking (or giving) additional income from (or to) group B and the marginal cost of doing so, which in turn depends on the value that group B places on additional income in a given state. An optimal trade-off involves efficient risk sharing, in the sense that it is optimal for group A to provide full insurance to group B against fluctuations in the insurable component of their pre-transfer incomes.¹¹

¹¹An efficient social contract specifies a schedule of transfers that is independent of the realization y_A ,

To see why Pareto efficient paths must be stationary, note that the above efficiency condition has implications for consumption smoothing across time periods. Intuitively, efficient consumption smoothing reflects the equality of the two groups' marginal rates of substitution between current and next-period's consumption, that is,

$$\frac{E(u'(y_B^t + T^t))}{\delta E(u'(y_B^{t+1} + T^{t+1}))} = \frac{1}{\delta}, \quad (11)$$

which implies that $T^{t+1} = T^t$ for all t , since incomes are i.i.d., all individuals have the same discount factor, and individuals in group B have a strictly decreasing marginal utility.

The next proposition shows that any interior Pareto efficient path can be implemented by a subgame perfect equilibrium if individuals are sufficiently patient. It also characterizes efficient perfect equilibria in terms of strategy profiles that are *simple* in the sense of Abreu (1988). Formally, a *simple strategy profile* $\sigma(Q_0, Q_A, Q_B)$ is a rule that specifies: (i) play Q_0 until some group deviates singly from Q_0 ; (ii) for $i = A, B$, play Q_i if group i deviates singly from Q_j , for $j = 0, A, B$, where Q_j is an ongoing previously specified path; continue with Q_j if no deviations occur or if both groups deviate simultaneously.

Simple strategy profiles are history-independent in the sense that they specify the same punishment Q_i for any deviation by player i after any previous history. Also note that group A chooses transfers for every realization of the insurable/taxable components of income. Thus, a deviation by group A is a deviation from the contract that applies to all insurable/taxable income realizations, where we have assumed that all deviations from the equilibrium play are observable.

Abreu's analysis applies here, and it implies that $\sigma(Q_0, Q_A, Q_B)$ is perfect if and only if no one-shot deviation by any group $i = A, B$ from Q_j , $j = 0, A, B$, is profitable, given that both groups will conform with Q_i after the deviation. It also implies that a path Q_0 is the outcome of a subgame perfect equilibrium (σ_A, σ_B) if and only if it is the outcome of some perfect simple strategy profile σ .

Proposition 2 constructs perfect simple strategy profiles that can implement any given interior Pareto efficient path, provided that individuals are sufficiently patient. Interior paths

because group A 's members are risk neutral.

are paths where each group's payoffs are strictly individually rational every period. Since Pareto efficient paths are stationary, they are interior if and only if $U_B(Q) \in \text{int } V_B$.

Proposition 2 *For any interior Pareto efficient path $Q_e = \{(1, T_e)\}_{t \geq 0}$, there is a discount factor $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$ there is a perfect simple strategy profile $\sigma^*(Q_e, Q_A, Q_B)$ that implements Q_e , where any unilateral deviation by group A (from any ongoing path) triggers a switch to the path Q_A involving $\tau \in (0, \infty)$ periods of conflict followed by a return to the equilibrium path and any unilateral deviation by group B (from any ongoing path) triggers a switch to the path Q_B in which the Nash equilibrium of the stage game is played permanently.*

Formally, the proof of Proposition 2 in the Appendix shows that $\sigma^*(Q_e, Q_A, Q_B)$ can implement Q_e by specifying the paths (punishments):

$$Q_A = \left\{ \{(0, T_{NE})\}_{t=0}^\tau, \{(1, T_e)\}_{t>\tau} \right\}, \text{ for some } \tau \in (0, \infty), \quad (12)$$

$$Q_B = \{(1, T_{NE})\}_{t \geq 0}, \quad (13)$$

where T_e is given by Proposition 1 and T_{NE} is given by equation (6). Not surprisingly, subgame perfection requires the punishment Q_A to take the form of a stick-and-carrot punishment, where the earlier stages of the punishment are more unpleasant than the remainder. In particular, note that the group B's threat of opting out of the social contract forever (i.e., $\tau = \infty$) would not be a credible threat.

Proposition 2 characterizes a particular structure of behavior underlying social contracts that provide efficient risk sharing. Such a structure may be interpreted in terms of a social norm involving a credible threat of conflict by the non-governing group that serves to constrain the power of the governing group to influence the distribution of the social surplus through social policy. Proposition 1 and Proposition 2 together imply that how the surplus is actually distributed is a function of the equilibrium. Thus, different equilibria can be interpreted as differing in terms of the (endogenous) bargaining power of the two groups. While the source of such a bargaining power is left unexplained, it is natural to view it as

the result of the underlying institutional environment.¹²

Moreover, any (interior) efficient perfect equilibrium implements the social contract that solves the following static problem every period:

$$\max_{T(y_A, \{y_j\}_{j \in K})} \{E(y_A^t - T)\} \text{ subject to } E(u(y_B^t + T)) \geq v \in \text{int } V_B. \quad (14)$$

Conversely, any solution to the above problem can be implemented by an efficient perfect equilibrium, if individuals are sufficiently patient. In this context, one can interpret σ^* as a rule that specifies equilibrium *social demands*. Group A demands that group B participates in the social contract, while the latter demands social policy (i.e., a social contract). One can also think of social demands by group B every period in terms of utility (v) rather than transfers (T). Group A can then be viewed as the residual claimant to the social surplus every period after meeting group B 's "utility demands". It is therefore in the self-interest of group A to design policies that meet utility demands in the least costly manner. This entails meeting the participation constraint of group B by providing insurance, in the form of state-contingent transfers, rather than pure redistribution.

4 Top income taxation and social insurance

In this section we extend our basic model of social policy in order to formalize how the Great Depression, and how the proliferation of social insurance programs, may have each influenced the taxation of top incomes. To that end, first note that Proposition 1 implies that the expected tax rate paid by the governing group in the basic model is simply

$$E\left(\frac{1}{y_A^t} T_e\right) = E\left(\frac{1}{y_A^t}\right) E\left(c - \sum_{j \in K} a_j y_j^t\right), \quad (15)$$

where the equality follows from the fact that T_e and y_A are independent random variables. The taxes paid by group A are the transfers received by group B . Group A will be a net

¹²Assuming that the governing group has an outside option (e.g., allowing for capital flight) would generally influence the set of interior efficient paths, but it would not change our main results, provided that the outside option involves an inefficient outcome.

payer for all income realizations if group B 's utility demands (v) are high enough. We want to analyze how the expected tax rate $E(T_e/y_A^t)$ responds first to changes in aggregate risk and second to changes in insurance opportunities.

The level of the expected tax $E(T_e)$ is a function of the distribution of the insurable income components $\{y_A^t, \{y_j^t\}_{j \in K}\}$ only through $E\left(\sum_{j \in K} a_j y_j^t\right)$, but it is a more complicated function of the distribution of the uninsurable components $\{y_j^t\}_{j \notin K}$. An increase in aggregate risk can be understood in terms of a mean-preserving spread in the distribution of the sum of uninsurable risks $\sum_{j \notin K} a_j y_j^t$. Since we have assumed that all risky components are independent and identically distributed we can formalize an increase in aggregate risk as a mean-preserving spread in the common distribution $F(y_j^t)$.

Since the increase in uninsurable risk does not affect the distribution of the governing group's pre-tax income, all we need to know is its effect on the expected tax $E(T_e)$.¹³ The increased riskiness of the insurable components of risk has no effect on the expected transfers, because the governing group is risk neutral.

Now consider changes in insurance opportunities. We have introduced multiple risks to capture a notion of different public insurance programs, each targeting a specific risk. Thus, we interpret the social contract T_e as relying on a single social insurance program for each insurable risk $j \in K$, and thus, we can view the number of social insurance programs as being identical to the number of insurable risks $j \in K$. From this perspective, one can formalize the effect of introducing additional social insurance programs as an expansion in the subset of insurable risks from K to K' , with $K \subset K'$. The expansion of insurance programs does not change the governing-group's pre-transfer income distribution, and so, once again, all we need to know is its effect on the expected tax $E(T_e)$.

Simply considering the effect of exogenous changes in F or K in the context of the above stationary environment is somewhat problematic. For a given equilibrium, the unexpected change would have no impact on behavior, and moreover, the original social contract T_e would

¹³Alternatively, note that if the increase in aggregate risk were to be accompanied by a mean-preserving spread in the distribution of y_A^t , the effect on the expected tax rate through $E(1/y_A^t)$ directly would be positive, since $1/y_A^t$ is a convex function of y_A^t . As we will show, this effect would reinforce the effect of a mean-preserving spread in the distribution of aggregate uninsurable risk on the expected tax rate.

no longer be efficient after the change in the environment. If the original equilibrium does change in response to unexpected changes in the environment, the multiplicity of equilibria in the repeated game implies that equilibrium responses are arbitrary. We address this problem next, by treating the distribution of uninsurable risks and the set of insurable risks at time t , denoted by F^t and K^t , respectively, as state variables that evolve over time.

4.1 Changes in aggregate risk and in insurance programs

The new assumption is that now the distribution of uninsurable risks F^t and the subset of insurable risks K^t evolve as follows. In period 0 the state is given by (F_0, K_0) . In subsequent periods there is a constant probability $\varepsilon > 0$ every period that the distribution of uninsurable incomes will change from F_0 to F_1 and will remain F_1 forever after, where F_1 is a mean-preserving spread of F_0 . In turn, once the state (F_1, K_0) is reached, there is a constant probability $\pi > 0$ every period that the subset of insurable incomes will change from K_0 to K_1 , with $K_0 \subset K_1$, and will remain K_1 forever after. Accordingly, $(F^t, K^t) = (F_1, K_1)$, is an absorbing state. Thus, one needs to consider only two (stochastic) transitions: one from (F_0, K_0) to (F_1, K_0) and a second one from (F_1, K_0) to (F_1, K_1) . For simplicity, we assume that the distributions F_0 and F_1 have common support. We also assume that the state (F^t, K^t) is observed at the beginning of period t , before date- t actions are taken.

Our analysis of the basic model with a constant environment carries through with appropriate modifications, where one needs to account for the fact that histories of the game include the history of realizations of the exogenous state (F^t, K^t) . Let $\hat{h}^t = \{(x^\tau, T^\tau, Y^\tau, F^\tau, K^\tau)\}_{\tau=0}^{t-1}$ be a history of the game at date t . The game starts at date 0 with the null history \hat{h}^0 . A *pure strategy profile* $(\hat{\sigma}_A, \hat{\sigma}_B)$ is a sequence of functions $\hat{\sigma}_i = \left\{ \hat{\sigma}_i^t \left(\hat{h}^t, F^t, K^t \right) \right\}_{t \geq 0}$, for $i = A, B$.

Since histories are random, one must recognize the dependence of actions on the current state, as implied by $(\hat{\sigma}_A, \hat{\sigma}_B)$. Accordingly, let a path be a sequence $\hat{Q} = \left\{ \left(\hat{x}^t, \hat{T}^t \right) \right\}_{t \geq 0}$, where $\hat{x}^t(F^t, K^t)$ and $\hat{T}^t(y_A, \{y_j\}_{j \in K^t}, F^t, K^t)$ are functions specifying the actions x^τ and T^τ to be taken in every state. Let $\hat{\Omega}$ denote the set of feasible paths. Note that (\hat{x}^t, \hat{T}^t) specifies actions at date t for each state $(F, K) \in \{(F_0, K_0), (F_1, K_0), (F_1, K_1)\}$ and not just for the realized state at date t . Every strategy profile generates a path that can be defined

inductively in the obvious manner.

Preferences over paths for each individual in group A and B are now given by

$$\widehat{U}_A(\widehat{Q}) = (1 - \delta)E_0 \left(\sum_{t=0}^{\infty} \delta^t \widehat{x}^t E \left(y_A^t - \widehat{T}^t \mid F^t, K^t \right) \right), \quad (16)$$

and

$$\widehat{U}_B(\widehat{Q}) = (1 - \delta)E_0 \left(\sum_{t=0}^{\infty} \delta^t \widehat{x}^t E \left(u \left(y_B^t + \widehat{T}^t \right) \mid F^t, K^t \right) \right), \quad (17)$$

respectively, where E denotes the expectation with respect to income realizations conditional on the realized state (F^t, K^t) at date t , and E_0 denotes the expectation with respect to the distribution over future histories of the state.

The set of feasible individually rational payoffs in the stage game depends on time only through the current realization of the state, and it is given by

$$\widehat{V}_A(F, K) = \left[0, E \left(y_A^t + \sum_{j \in K^t} a_j y_j^t \mid F, K \right) \right], \quad (18)$$

and

$$\widehat{V}_B(F, K) = \left[E \left(u \left(\sum_{j \notin K^t} a_j y_j^t \right) \mid F, K \right), E \left(u \left(y_A^t + y_B^t \right) \mid F, K \right) \right], \quad (19)$$

for $(F, K) \in \{(F_0, K_0), (F_1, K_0), (F_1, K_1)\}$, for groups A and B , respectively, at any date t .

Proceeding as in Section 3.3, we now characterize interior efficient perfect equilibria. Proposition 3 describes the properties of interior Pareto efficient paths, and Proposition 4 shows how they can be implemented by subgame perfect equilibria.

Proposition 3 *An interior path $\widehat{Q} = \left\{ \left(\widehat{x}^t, \widehat{T}^t \right) \right\}_{t \geq 0}$ is Pareto efficient if and only if it has the following properties:*

- (1) *the non-governing group always participates ($\widehat{x}^t = 1$, for all (F, K) , and for all $t \geq 0$);*
- (2) *the social contract provides full insurance against insurable shocks every period by*

specifying the state-dependent transfer schedule $\widehat{T}^t = \widehat{T}_e$ for all $t \geq 0$, where

$$\widehat{T}_e = c(F, K) - \sum_{j \in K} a_j y_j, \text{ for } (F, K) = (F_0, K_0), (F_1, K_0), (F_1, K_1),$$

and $c(F_1, K_1) > 0$ is a constant satisfying

$$E \left(u \left(c(F_1, K_1) + \sum_{j \notin K^t} a_j y_j^t \right) \mid F_1, K_1 \right) = v \in \text{int } \widehat{V}_B(F_1, K_1);$$

(3) the social contract provides efficient consumption smoothing across states, that is, $c(F_0, K_0) > 0$ and $c(F_1, K_0) > 0$ are two constants such that expected marginal utilities (conditional on the state) for the non-governing group are equated across states.

It is not difficult to verify that the set of interior Pareto efficient paths is non-empty if the average income of the governing group, $E(y_A^t)$, is sufficiently large.

As in Section 3.3, the set of Pareto efficient paths can be characterized by maximizing group A 's utility, $\widehat{U}_A(\widehat{Q})$, over all paths subject to group B 's utility, $\widehat{U}_B(\widehat{Q})$, being no lower than a given value v , for each feasible value v . It is easy to verify that the solution to this problem is such that $\widehat{x}^t(F, K) = 1$, for all t and all (F, K) , and it satisfies

$$1 = \widehat{\lambda} E \left(u' \left(y_B^t + T^t \right) \mid y_A, \{y_j\}_{j \in K^t}, F, K \right), \quad (20)$$

for all $\{y_A, \{y_j\}_{j \in K}\}$, for all (F, K) , and for all $t \geq 0$, where $\widehat{\lambda}$ is the Lagrange multiplier for group B 's participation constraint.

Equation (20) has several important implications. It implies that every period the governing group provides full insurance to the non-governing group against fluctuations in the insurable component of their incomes. It also implies that efficient social contracts depend on time only through the state (F, K) . Accordingly, $T^{t+1} = T^t$ if $(F^{t+1}, K^{t+1}) = (F^t, K^t)$. Moreover, once the absorbing state (F_1, K_1) is reached, the analysis in Section 3 applies, and efficient social contracts become stationary. Then, for any arbitrary efficient contract in state (F_1, K_1) , the social contracts in the other two feasible states (F_0, K_0) and (F_1, K_0) are also implied by equation (20), which requires that expected marginal utilities (conditional

on the state) be equated across states:

$$E(u'(y_B^t + T^t) | F_0, K_0) = E(u'(y_B^t + T^t) | F_1, K_0) = E(u'(y_B^t + T^t) | F_1, K_1). \quad (21)$$

This condition reflects efficient consumption smoothing across states as well as time. Consequently, the characterization of efficient paths in Proposition 3 is independent of the probability distribution of changes in (F^t, K^t) , as given by the probabilities ε and π .¹⁴ This reflects the fact that the non-governing group is in effect insured against the risk of changes in the state (F^t, K^t) .

One can also construct perfect simple strategy profiles that implement any given interior Pareto efficient path, provided that individuals are sufficiently patient.

Proposition 4 *For any interior Pareto efficient path $\widehat{Q}_e = \left\{ (\widehat{x}_e, \widehat{T}_e) \right\}_{t \geq 0}$, there is a discount factor $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$ there is a perfect simple strategy profile $\widehat{\sigma}^* (\widehat{Q}_e, \widehat{Q}_A, \widehat{Q}_B)$ that implements \widehat{Q}_e , where any unilateral deviation by group A (from any ongoing path) triggers a switch to the path \widehat{Q}_A involving $\tau \in (0, \infty)$ periods of conflict followed by a return to the equilibrium path and any unilateral deviation by group B (from any ongoing path) triggers a switch to the path \widehat{Q}_B in which the Nash equilibrium of the stage game is played permanently.*

While the perfect strategy profile $\widehat{\sigma}^*$ described in Proposition 4 must specify actions (x^t, T^t) at any date t as a function of the state (F, K) , as dictated by the functions $(\widehat{x}_e, \widehat{T}_e)$, the structure of $\widehat{\sigma}^*$ parallels that of the perfect strategy profile σ^* characterized in Proposition 2 for the case of a constant state. Formally, $\widehat{\sigma}^* (\widehat{Q}_e, \widehat{Q}_A, \widehat{Q}_B)$ implements $\widehat{Q}_e = \left\{ (\widehat{x}_e, \widehat{T}_e) \right\}_{t \geq 0}$ by specifying the two paths (punishments)

$$\widehat{Q}_A = \left\{ \left\{ (\widehat{x}_c, \widehat{T}_{NE}) \right\}_{t=0}^{\tau}, \left\{ (\widehat{x}_e, \widehat{T}_e) \right\}_{t > \tau} \right\}, \text{ for some } \tau \in (0, \infty), \quad (22)$$

$$\widehat{Q}_B = \left\{ (\widehat{x}_{NE}, \widehat{T}_{NE}) \right\}_{t \geq 0}, \quad (23)$$

¹⁴Note that the probabilities ε and π do not affect the set of efficient paths, because we have normalized conflict payoffs to zero, and so there are always positive gains from cooperation.

where $(\widehat{x}_e, \widehat{T}_e)$ is given by Proposition 3, and

$$(\widehat{x}_c, \widehat{T}_{NE}) = \left(0, -\sum_{j \in K} a_j y_j \right), \text{ for } (F, K) = (F_0, K_0), (F_1, K_0), (F_1, K_1); \quad (24)$$

$$(\widehat{x}_{NE}, \widehat{T}_{NE}) = \left(1, -\sum_{j \in K} a_j y_j \right), \text{ for } (F, K) = (F_0, K_0), (F_1, K_0), (F_1, K_1). \quad (25)$$

In the next section we discuss the implications of the above efficient perfect equilibria.

4.2 Implications for top income taxation

In order to understand the implications of our analysis for top income taxation, it is useful to think of social demands by the non-governing group every period in terms of utility rather than transfers. To that end, consider an efficient perfect equilibrium $\widehat{\sigma}^* (\widehat{Q}_e, \widehat{Q}_A, \widehat{Q}_B)$, and note that $\widehat{\sigma}^*$ generates a unique sequence of payoffs $\{v_A(F^t, K^t), v_B(F^t, K^t)\}_{t \geq 0}$, where $v_A(F, K) = E(y_A^t - \widehat{T}_e | F, K)$ and $v_B(F, K) = E(u(y_B^t + \widehat{T}_e) | F, K)$ depend on time only through the state (F^t, K^t) .

Behavior along the equilibrium path can be viewed as a situation where the non-governing group demands $E(u(y_B^t + T^t) | F^t, K^t) \geq v_B(F^t, K^t)$ in period t , and the governing group supplies an optimal social contract that meets the non-governing group's participation constraint that period. Formally, Proposition 3 and Proposition 4, together, imply that any interior efficient perfect equilibrium generating payoffs $v_A(F, K)$ and $v_B(F, K)$ in state (F, K) implements the social contract that solves the following static problem every period, as a function of the realized state (F, K) :

$$\max_{T(y_A, \{y_j\}_{j \in K}, F, K)} \{E(y_A^t - T | F, K)\} \text{ subject to } E(u(y_B^t + T) | F, K) \geq v_B(F, K). \quad (26)$$

Conversely, any solution to problem (26), for $(F, K) = (F_0, K_0), (F_1, K_0), (F_1, K_1)$, such that expected marginal utilities for the non-governing group are equated across states (i.e., such that equation (21) holds) can be implemented by an efficient perfect equilibrium, if individuals are sufficiently patient. In particular, note that Proposition 3 characterizes the value of the payoffs $v_B(F_0, K_0)$ and $v_B(F_1, K_0)$ as a function of the payoff $v_B(F_1, K_1) = v$

for all interior Pareto efficient paths.

Even though the state (F, K) changes over time, the governing group can still be viewed as the residual claimant to the social surplus every period after meeting the utility demands of the non-governing group. It is therefore in the self-interest of the governing group to design policies that meet those utility demands every period in the least costly manner, which is with social insurance.

Proposition 5 highlights the role of aggregate risk, and the connection between social insurance and top income taxation, which we think sheds new light into the common temporal pattern of modern welfare states. For instance, while others have viewed the Great Depression as a pivotal event, the relevant mechanism is not apparent. If the Depression led to increased demands for redistribution, why were social programs aimed at providing insurance rather than pure redistribution? If it increased aggregate risk, what is the connection between aggregate risk, insurance, and top income taxation?

Proposition 5 *Along any interior Pareto efficient path $\hat{Q}_e = \{\hat{x}_e, \hat{T}_e\}_{t \geq 0}$ expected top income taxes increase with aggregate risk, that is, $E(\hat{T}_e | F_1, K_0) > E(\hat{T}_e | F_0, K_0)$, and they decrease with the introduction of additional insurance programs, that is, $E(\hat{T}_e | F_1, K_1) < E(\hat{T}_e | F_1, K_0)$.*

Both results stated in the proposition are driven by a common mechanism: social insurance and top income taxation are substitutes for meeting the non-governing group's utility demands along an efficient (equilibrium) path, hence for avoiding social conflict.

To see this, consider a given efficient path, where initially group B enjoys a given payoff $v_B(F_0, K_0)$ every period. Now consider an arbitrary period t where aggregate risk increases, as described by the change from F_0 to F_1 , where F_1 is a mean-preserving spread of F_0 . The response of the expected tax paid by group A depends on how the expected *marginal utility* of consumption for individuals in group B responds to the increase in aggregate risk. Since marginal utility is convex (that is, $u''' > 0$), such an increase tends to raise expected marginal utility in period t . In order to keep marginal utility constant, average transfers to group B must increase so as to increase their average consumption. This is the natural effect

of efficient risk sharing, whereby group B in effect demands a higher average consumption when individuals in the group experience larger levels of risk. Note that the effect on the level of group B 's expected utility is in general ambiguous.

Now suppose that new insurance programs arrive at date t . If the social contract did not change, consumption smoothing would still be efficient, but insurance would not. Accordingly, it is in the interest of the governing group to take advantage of the new insurance opportunities in order to reduce the risk faced by the non-governing group. In turn, the reduction in risk tends to lower expected *marginal utility* in period t ; therefore group A can keep group B 's expected marginal utility constant by lowering average transfers to group B so as to decrease their average consumption. Once again, there is a trade-off between insurance and top income taxation, but in this case group A exploits the trade-off to lower the average tax they need to pay to ensure that group B participates in the social contract.

Further remarks

It is not difficult to extend our previous analysis to the case of multiple non-governing groups. The nature of efficient perfect equilibria would remain unchanged under the assumption that the governing group is able to condition transfers on *group identity* as well as on income realizations. Intuitively, the model could then capture the idea that social groups can hurt group A relatively more when they are able to coordinate their opting out decisions. Now the background where social demands arise is one in which the degree of social coordination among the non-governing groups and the ability of the governing group to prevent such coordination play a critical role in the allocation of the social surplus. From this perspective, more redistributive welfare states result when members of a society can overcome their coordination problems and agree to place larger social demands on the state. Hence, it is not necessarily the case that the threat of social unrest was historically larger in countries such as Sweden, but that social demands were. In turn, one would expect that the ability to coordinate social demands depends largely on political institutions, and that social coordination may be easier in more homogeneous countries.

When there are multiple non-governing groups, an implication of perfect efficient equi-

libria is that, whenever feasible, it is in the best interest of the governing group to design distinct *group-specific* transfer schedules, each restricting eligibility to individuals in groups facing common risks (i.e. the same profile of a_j 's). At any given time, the governing group may not have enough instruments to achieve this outcome, which is a source of inefficiency. For instance, some transfers may be conditioned only on observable group characteristics, such as race, gender, union membership or old age, which can serve as proxies for socioeconomic groups of individuals with similar risk characteristics and/or social demands. Other transfers may be conditional on income alone, without respect for the differences in the actual risk faced by different individuals. This too is a source of inefficiency. The governing group becomes better off as these constraints are lifted, because it can meet any given utility demands at a lower cost for itself.

The previous argument is interesting because it can explain why the growth of the welfare state has consisted of an enlargement of insurable events characterized by the increasing number and complexity of transfer programs as well as the narrowing target of each program. Our model suggests a view of the growth of the welfare state as being driven by the lifting of technical constraints faced by the public administration.

5 Social policy in historical perspective

In this section, we argue that the political influence of the wealthy, the threat of social conflict, and the targeting of heterogeneous risks, all played major roles in the evolution of social policy over the 20th century. Our most detailed arguments are for the United States, but we also discuss the cases of Britain and Sweden.

United States

Despite four episodes of high unemployment in the United States from 1890 until just prior to the Great Depression, no American state had passed legislation for unemployment insurance, old-age insurance, or health insurance before 1929. One possible reason for the muted effect of the 1890s depression on social insurance reform is that there was a lack of common interests among American workers in the late 19th century: in terms of our model, an absence of

coordination among non-governing groups. In comparing the depression of the 1890s with the Great Depression of the 1930s, Rauchway (2008: 38) notes: “the 1890s fell during an age of globalization, ... so many of the country’s workers had been immigrants, literally of another people. By the Great Depression, this was no longer true.” Furthermore, the massive unemployment during the Great Depression cut across social classes and bankrupted many trade union pension plans (Weaver, 1983). These observations suggest why social protest escalated during the Great Depression.¹⁵

The Social Security Act of 1935 marks the birth of the welfare state in the United States. It provided federal benefits to retirees and grants to states for the unemployed, old-age assistance, aid to dependent children, vocational rehabilitation, maternal and child welfare, and public health work.¹⁶ The New Deal programs were legislated as social insurance to combat economic insecurity, with an emphasis on the fact that “funds for the payment of insurance benefits should not come from the proceeds of general taxation.”¹⁷ However, they were accompanied by sharp rises in tax progressivity.¹⁸ Our interpretation is that the need for redistribution as an accompaniment to social insurance to avert social conflict was recognized at the time. Thus, while advocating the introduction of an inheritance tax, President Roosevelt stated in a message to Congress on Tax Revision on June 19, 1935: “Social unrest and a deepening sense of unfairness are dangers to our national life which we must minimize by vigorous methods.” He noted further that, “The movement toward progressive taxation of wealth and of income has accompanied the growing diversification and interrelation of effort which marks our industrial society.” These comments point to the use of progressive taxation for securing industrial cooperation during the Great Depression.

Domhoff (1990, 1996) discusses the identity of the power elite and the channels through which they exert their power, while Dal Bó et al. (2009) provide evidence of the remarkable

¹⁵Kerbo and Shaffer (1986) estimated the number of unemployed protestors in the key years of recession/depression in the U.S. before 1929. The numbers of protestors were: 500 in 1894; 4,800 in 1908; 3,500 in 1914; and 450 in 1922. In contrast to this earlier period, there were 102,500 protestors in 1930 and the annual average number of protestors from 1930-1940 was 49,479.

¹⁶Prior to 1935 social assistance in the U.S. consisted of minor programs from state and local governments and self-help trade associations.

¹⁷President Roosevelt’s message to Congress on Social Security, January 17, 1935.

¹⁸For instance, the top marginal tax rate on personal income went from 24 percent in 1929 to 63 percent in 1932 and 79 percent in 1938.

persistence of political dynasties in the United States.¹⁹ The important role of elites in the adoption of the 1935 Act has been stressed by Domhoff and also by Quadagno (1984: 645), who remarks that “[I]t is also important to explain why a piece of legislation with such a high level of ‘class content’, i.e., a social-welfare measure, was implemented with almost no working-class input.”²⁰

Alston and Ferrie (1993) observe that Southern large plantation owners formed a major segment of the U.S. governing group for a century preceding 1970, operating through overwhelming Congressional committee control by the Democrats. The rural elite’s political objectives were to maintain low wages and to secure federal agricultural subsidies for large plantations. The Southern landowners ensured that the unemployment insurance component of the 1935 Act excluded farm workers and that public assistance programs were administered by the states rather than the federal government. Alston and Ferrie also note that farm mechanization in the 1950s significantly reduced the need for plantation workers and that it was increasingly in the interest of Southern landowners to promote the out-migration of blacks to Northern states. Thus, commenting on the Economic Opportunity Act, which was the centerpiece of the Johnson Administration’s War on Poverty, they argue that “it is extremely unlikely that the welfare state programs of the 1960’s could have emerged from Congress without the countenance of Southern Congressmen”.

The development of the welfare state since the Great Depression has been one of enlargement of insurable events and gradual extensions of eligibility to include more groups of individuals. The original Social Security Act excluded numerous job categories encompassing about half the working population. Coverage was added for various workers in subsequent decades. The 1950 Social Security amendments extended the old-age and survivors insurance programs to about 10 million more persons. Self-employed farmers, homeworkers, and some self-employed professionals were added in 1954. The inclusion of disability benefits into the Social Security Act in 1956 is an example of an increase in transfers resulting from

¹⁹For example, the Breckinridge family placed 17 members in Congress, serving 72 congresses, from 1789 to 1978.

²⁰It is also noteworthy that large manufacturers demanded legislation for social insurance during the formative years of the welfare states in France and Germany (Mares, 2003).

an enlargement of insurable events. The 1960 Social Security amendments increased federal grants to states to provide medical care for aged people receiving social assistance. Medicare and Medicaid were established in 1965. Old Age Insurance paid benefits to 16 percent of the aged in 1950, but to 92 percent by 1978 (Lampman, 1984).²¹

The total tax burden imposed by federal taxation has been approximately constant as a proportion of GDP from 1953 onwards, but the composition has changed substantially.²² Social insurance contributions have risen steadily since 1950 while corporation income taxes have fallen by a similar magnitude beginning in 1953 (see Table 2.3 in the Budget of the United States, 2005). Since payroll taxes are regressive and the corporation income tax is progressive, the replacement of corporate taxes with payroll taxes in federal receipts contributed to the decline in tax progressivity. Piketty and Saez (2007) find that the average federal tax rate, including individual, corporate, payroll, and estate taxes, for the top 0.01 percent of the income distribution declined from 71.4 percent in 1960 to 59.3 in 1980, whereas corresponding rates faced by income groups between the 40th and the 95th percentiles were higher in 1980 than they were in 1960.

Our interpretation of this decline in top income taxation is that a gradual lifting of supply-side constraints enabled more comprehensive social insurance coverage and, consequently, reduced the need for redistribution. Learning-by-doing and technical improvements in the administration of programs and taxes played an important role in expanding coverage and lowering the cost of tax compliance. The payroll tax in the Social Security Act, for example, was a precursor to the general withholding tax on wages introduced in 1943. Quoting from U.S. Senate Hearings (July-August 1942: 136): “All of the employers have had 7 years’ experience [and] they will be in a much better position to do this job than they were to do the social security job back in 1936 when it first went into effect.”²³ Furthermore, advances in record-keeping technology, especially automated data processing that originated during

²¹See Kollman (1996) for the major legislative changes in social security benefits and coverage by employment categories.

²²Throughout the post-WWII period until the mid-1970s, budget deficits were modest, and debt-to-GDP ratios were declining, not just in the US, but across the industrial economies. In the US, federal spending as a share of GDP remained stable between 1953 and 1980, as the increase in social transfers was offset by a relative decline in other federal spending, most notably military expenditures.

²³Cited in Twight (1995).

World War II, likely facilitated the extensions of social insurance programs in the United States in the 1960s and the 1970s.²⁴

Great Britain

Social legislation was not prominent in local elections between 1885 and 1914 and voter turnout was low in the United Kingdom (Thane, 1984). However, social unrest during this period is apparent in the strike waves occurring in 1889 and the years immediately following and especially in 1910 (Cronin, 1979). In this setting the Old Age Pension Act of 1908 and the National Insurance Act of 1911, which covered unemployment and health, were introduced by Winston Churchill and Lloyd George in the Liberal government led by Herbert Asquith. The initiatives can be viewed as preemptive attempts to improve industrial relations. According to Steinmo (1993):

“The two programs [health and unemployment insurance], conceived and designed by relatively insulated groups of party officials and civil servants, were not directly the result of pressure from the Left, but rather stemmed from the government’s attempts to head off growing support for socialism generally and the Labour party specifically.”

Unemployment insurance coverage was extended to include most manual workers in 1920, although the size of transfer payments remained small until the Great Depression. In 1931 the scheme was replaced by full government funding, whereby benefits were tied to need rather than to contributions. The latent threat of unrest in the inter-war period in the United Kingdom is reflected in the General Strike of 1926 and in two major phases of unemployed agitation organized by the National Unemployed Workers’ Movement (NUWM) from 1919–22 and 1931–36. In early 1932, thousands of unemployed battled with police in London and Bristol and 100,000 demonstrated in Newcastle (Perry, 2000).

Unrest in the inter-war period did not by itself produce significant changes to programs, but it is part of the background for the comprehensive social insurance reforms proposed

²⁴In 1956 the Social Security Administration installed the first large-scale computer to maintain records and an integrated data processing system was put into effect in 1965.

by the Beveridge Report in 1942 as a response to the experience of the Great Depression. The Report identified six classes of vulnerable populations: employees, the self-employed, housewives, others of working age but not gainfully employed, those below working age, and the retired above working age. Implementation of the Report included the Family Allowances Act 1945, the National Insurance Act 1946, and the National Health Service Act 1946.

The formation of British welfare policy was influenced by business elites partly through the primacy given to the Treasury department's concerns and through the requirement that senior civil servants be educated at either Oxford or Cambridge in order to be exposed to the priorities of Treasury. The alignment of Treasury's outlook with elite interests resides in the importance of London as a trade and financial centre and in Treasury's role as the coordinating department of the civil service since 1868 (Valocchi, 1989). The cost of the Beveridge Committee's initial proposals in 1942 caused considerable concern in the Treasury, after which Beveridge "made some major concessions (and thereby partly jeopardised the Report's principles) in an attempt to cut costs" (Lowe, 1999: 135). Treasury continued its campaign to reduce the role of taxation in financing pensions over the next two decades, regardless of the party in power, which gradually reduced the redistributive element of the postwar pension system (Bridgen, 2006: 18).

The top marginal personal income tax rate in the U.K. rose from 50 percent in 1928 to 66 percent by 1931, and remained over 95 percent between 1941 and 1954 before falling in steps to 40 percent in 1988. Part of the reversal in tax progressivity occurred due to changes in the structure of taxation. Thus the share of total tax revenues raised from companies was halved between 1955 and 1964, and again between 1965 and 1974, while social security contributions and indirect taxes increased to compensate for the revenue shortfalls (Lowe, 1999: 286).

The introduction of the Pay-As-You-Earn method of payroll tax deductions in the U.K. in 1944 was a significant tax innovation, which enabled spending growth after World War II. Subsequently, programs were tailored to vulnerable groups and became means-tested: in 1975 there were 45 major means-tested benefits, with differing assessment criteria (Lowe, 1999: 152). Enhancements in social insurance coverage included changes to pensions for

widowers over the age of 50 in 1956, Supplementary Benefits in 1966, invalidity pensions in 1971, and an extension of earnings-related pensions to uncovered occupations in 1975 (Flora, 1983). Like in the case of the United States, our interpretation is that the incrementalism apparent in the development of social insurance programs in the United Kingdom reflects gradual efficiency-enhancing adaptations of programs to target needs more finely.

Sweden

Sweden was also affected by the severe global depression in the 1890s, and the possibility of social insurance schemes was considered there too. The administrative governor of the Stockholm area established an investigation of unemployment insurance for Stockholm and other major cities in 1894, but “[T]he committee concluded that the topic was too new and that no definite recommendations could yet be made” (Hecló, 1974: 71).

A compulsory old age pension system was introduced in 1913 and a means-tested local unemployment benefits program in 1914. These were rudimentary insurance programs with very low benefits (Lundberg and Åmark, 2001). Together with progressive taxes on income, wealth, death, and corporations, the policies were a response of the ruling elite to working-class discontent (Steinmo, 1993). The elections in the 1920s were the first with the potential to reveal the political effects of universal suffrage achieved in 1921.²⁵ Yet, according to Hecló (1974: p. 290) the new voting rolls had no impact on social policy.

The origin of the modern welfare state in Sweden can be traced to an historic agreement between the Swedish Employers Federation (SAF) and the major trades union congress (LO) in May of 1936. Steinmo (1993: 88) argues that the postwar social policies of Sweden reflected this historic compromise between the agenda setting power of “big business” interests and the power of labor unions to disrupt production. This event was prompted by heightened labor unrest in the Great Depression. Workdays lost because of labor stoppages decreased almost four-fold between the periods of 1931–35 and 1936–40 (Steinmo, 1993). Regular semisecret meetings between leaders of the Social Democratic party and the SAF occurred beginning in the early 1950s. At these so-called Thursday club meetings, government and business elites

²⁵Near universal male suffrage was granted in 1909 but the greatest extensions in the voting rolls happened after World War I (Hecló, 1974).

discussed economic policy issues. Later labour officials from LO were added to the meetings. According to Steinmo (1993: 125), “many of the most difficult and controversial issues of the day were, in fact, settled behind closed doors by unelected representatives of interest organizations and technocrats.” Extraparliamentary politics remained important in Sweden at least until the mid-1970s.

Old age pensions were improved in 1935 and 1937 and became universal and state-financed in 1948, followed by accident insurance in 1954 and sickness insurance in 1955. The top marginal income tax rate increased from 39 percent in 1930 to 50 percent in 1934. The rate remained constant at 65 percent between 1953 and 1970, but tax progressivity was reduced by the introduction of a general sales tax at four percent in 1960, which was replaced by a value added tax at 10 percent in 1970. Furthermore, while the shares of personal income taxes, corporate taxes, and other taxes, in total receipts each fell from 1965 to 1985, the share of social security contributions more than doubled (OECD, 2008). Similarly to the U.S. and the U.K., changes in Swedish legislation led to a decline in the progressivity of income taxes between 1975 and 1985 (Bjorklund et al., 1995). Our interpretation of these events is that even the famously redistributive Swedish welfare state is primarily a social insurance arrangement, and that a decline in tax progressivity in Sweden ensued the expansion of social insurance.

6 Conclusions

The advice given to the Prince of Salina in *The Leopard* — Giuseppe di Lampedusa’s famous portrayal of socio-political upheaval in nineteenth century Italy — is: “If we want things to stay as they are, things will have to change.” In the same spirit, we have argued that, in the United States, Britain and Sweden, the wealthy responded to the threat of unrest brought about by the Great Depression by bearing a remarkable increase in top income taxation in the period 1930-1960. But as new social insurance programs and extensions of existing programs targeted heterogeneous risks more finely over time, less redistribution through the tax system was required to avert social unrest. By 1990, programs covered

the major categories of risk faced by most households and transfers stabilized. Our main conclusion is that the development of the welfare state reflected the gradual substitution of social insurance for top income taxation as an efficient means of averting social conflict.

We have abstracted from the roles commonly assigned to economic development and negative incentive effects in explaining the sources and the limits of the welfare state. We have shut down these effects in order to focus on the supply of social insurance and the role of economic risks. This reflects an important contrast between our view and one that equates a larger welfare state with larger efficiency losses. Instead, we see the growth in the number and complexity of transfer programs as reflecting the exploitation of further efficiency gains.

Our analysis has taken as given the pre-tax income distribution in order to focus on social policy. Conversely, analyses of the dynamics of the income distribution, particularly the top incomes, usually take social policy as given. We think that there are important insights to be gained from a better understanding of the joint evolution of the income distribution and social policy.

Appendix

Proof of Proposition 1

The set of Pareto efficient paths is characterized by the solution to the problem stated in equation (9). Clearly, Pareto efficient paths must be such that $x^t = 1$, for all $t \geq 0$, because members of group B receive at least their (positive) nontaxable income if they participate and zero otherwise, while members of group A can enjoy no less than their conflict payoff when group B participates. The discussion in the main text, following Proposition 1, shows that any such solution satisfies equation (10) for all $t \geq 0$, which implies that $T^{t+1} = T^t = c - \sum_{j \in K} a_j y_j^t$, for all $t \geq 0$. The proposition then follows from noting that $y_B^t + T_e = c + \sum_{j \notin K} a_j y_j^t$ and $E(u(y_B^t + T_e)) = U_B(Q)$, for all $t \geq 0$. **QED**

Proof of Proposition 2

Consider $\sigma^*(Q_e, Q_A, Q_B)$, where Q_e is an interior efficient path and Q_A and Q_B are given by equations (12) and (13). To verify that σ^* is perfect, it is sufficient to verify that no one-shot deviation by any group $i = A, B$ from Q_j , $j = e, A, B$, is profitable, given that both groups will conform with Q_i after the deviation. First note that group B has no incentive to deviate from Q_e , for any $\delta \in (0, 1)$, because $U_B(Q_B) < U_B(Q_e)$, for all $U_B(Q_e) \in \text{int } V_B$. It is also easy to verify that, for any given Q_e , there is a large enough $\tau < \infty$ such that group A has no incentive to deviate from Q_e , if $\delta < 1$ is sufficiently close to 1.

It remains to verify that no group has an incentive to deviate from either Q_A or Q_B once they are in force. For any $\delta \in (0, 1)$, neither group would have an incentive to deviate from Q_B because $(1, T_{NE})$ is a Nash equilibrium of the stage game. Moreover, it is now easy to see that group A would have no incentive to deviate from Q_A since then the punishment Q_A would be restarted. Finally, note that if group B conforms with Q_A each individual in the group enjoys utility $\delta^\tau U_B(Q_e)$, while they would enjoy $U_B(Q_B)$ if they were to deviate. Clearly, for any given $\tau \in (0, \infty)$, no deviation from Q_A by group B is profitable if $\delta < 1$ is sufficiently close to 1, because $U_B(Q_B) < U_B(Q_e)$, for all $U_B(Q_e) \in \text{int } V_B$. This concludes the proof. **QED**

Proof of Proposition 3

The proof parallels that of Proposition 1 and the main arguments are discussed following Proposition 3 in the main text, which imply that $c(F_0, K_0) > 0$ and $c(F_1, K_0) > 0$ are given

by

$$\begin{aligned} E \left(u' \left(c(F_0, K_0) + \sum_{j \notin K^t} a_j y_j^t \right) | F_0, K_0 \right) &= E \left(u' \left(c(F_1, K_0) + \sum_{j \notin K^t} a_j y_j^t \right) | F_1, K_0 \right) \\ &= E \left(u' \left(c(F_1, K_1) + \sum_{j \notin K^t} a_j y_j^t \right) | F_1, K_1 \right), \end{aligned}$$

as required. **QED**

Proof of Proposition 4

Consider $\hat{\sigma}^* (\hat{Q}_e, \hat{Q}_A, \hat{Q}_B)$, where \hat{Q}_e is an interior efficient path and \hat{Q}_A and \hat{Q}_B are given by equations (22) and (23). Letting $W_i^l(F, K)$ denote group i 's utility along the path \hat{Q}_l , for $l = e, A, B$, when the state is (F, K) , we have:

$$\begin{aligned} W_i^e(F_0, K_0) &= (1 - \delta) v_i(F_0, K_0) + \delta (\varepsilon W_i^e(F_1, K_0) + (1 - \varepsilon) W_i^e(F_0, K_0)), \\ W_i^e(F_1, K_0) &= (1 - \delta) v_i(F_1, K_0) + \delta (\pi W_i^e(F_1, K_1) + (1 - \pi) W_i^e(F_1, K_0)), \\ W_i^e(F_1, K_1) &= v_i(F_1, K_1), \end{aligned}$$

where $v_A(F, K) = E(y_A^t - \hat{T}_e | F, K)$ and $v_B(F, K) = E(u(y_B^t + \hat{T}_e) | F, K)$, and also:

$$\begin{aligned} W_i^A(F, K) &= \delta^\tau E_0(W_i^e(F^{\tau+1}, K^{\tau+1})) \\ W_i^B(F, K) &= \begin{cases} E \left(y_A^t + \sum_{j \in K^t} a_j y_j^t | F, K \right) & \text{if } i = A \\ E \left(u \left(\sum_{j \notin K^t} a_j y_j^t \right) | F, K \right) & \text{if } i = B. \end{cases} \end{aligned}$$

It is straightforward to verify that, for each state (F, K) , no one-shot deviation by any group $i = A, B$ from \hat{Q}_j , $j = e, A, B$, is profitable, given that both groups will conform with \hat{Q}_i after the deviation, if the discount factor is sufficiently high. For each state, the arguments parallel those used in the proof of Proposition 2. **QED**

Proof of Proposition 5

Consider an interior Pareto efficient path $\widehat{Q}_e = \{\widehat{x}_e, \widehat{T}_e\}_{t \geq 0}$. To prove that $E(\widehat{T}_e | F_1, K_0) > E(\widehat{T}_e | F_0, K_0)$, first note that $E(\widehat{T}_e | F, K) = c(F, K) - \sum_{j \in K} a_j E(y_j^t | F, K)$, for $(F, K) = (F_0, K_0), (F_1, K_0)$. Thus, $E(\widehat{T}_e | F_1, K_0) - E(\widehat{T}_e | F_0, K_0) = c(F_1, K_0) - c(F_0, K_0)$, since F_0 and F_1 have the same mean. It remains to prove that $c(F_1, K_0) > c(F_0, K_0)$. Note that

$$\begin{aligned} E\left(u'\left(c(F_1, K_0) + \sum_{j \notin K^t} a_j y_j^t\right) | F_1, K_0\right) &= E\left(u'\left(c(F_0, K_0) + \sum_{j \notin K^t} a_j y_j^t\right) | F_0, K_0\right) \\ &< E\left(u'\left(c(F_0, K_0) + \sum_{j \notin K^t} a_j y_j^t\right) | F_1, K_0\right), \end{aligned}$$

where the equality follows from Proposition 3, and the inequality follows from the facts that F_1 is a mean-preserving spread of F_0 , and u' is a strictly convex function (i.e., $u''' > 0$). It then follows that $c(F_1, K_0) > c(F_0, K_0)$, since u' is a strictly decreasing function.

To prove that $E(\widehat{T}_e | F_1, K_1) < E(\widehat{T}_e | F_1, K_0)$, note that

$$\begin{aligned} E\left(u'\left(c(F_1, K_1) + \sum_{j \notin K^t} a_j y_j^t\right) | F_1, K_1\right) &= E\left(u'\left(c(F_1, K_0) + \sum_{j \notin K^t} a_j y_j^t\right) | F_1, K_0\right) \\ &> E\left(u'\left(c(F_1, K_0) + \sum_{j \notin K_1} a_j y_j^t + \sum_{\substack{j \notin K_0 \\ j \in K_1}} a_j E(y_j^t)\right) | F_1, K_0\right) \end{aligned}$$

where the equality follows from Proposition 3, and the inequality follows from the facts that $K_0 \subset K_1$, with $\sum_{j \notin K_0} a_j y_j^t = \sum_{j \notin K_1} a_j y_j^t + \sum_{\substack{j \notin K_0 \\ j \in K_1}} a_j y_j^t$, and u' is a strictly convex function (i.e., $u''' > 0$). It then follows that

$$c(F_1, K_1) < c(F_1, K_0) + \sum_{\substack{j \notin K_0 \\ j \in K_1}} a_j E(y_j^t), \quad (27)$$

since u' is a strictly decreasing function (i.e., $u'' < 0$). Noting that $E(\widehat{T}_e | F, K) = c(F, K) -$

$\sum_{j \in K} a_j E(y_j^t | F, K)$, for $(F, K) = (F_1, K_0), (F_1, K_1)$, equation (27) implies

$$\begin{aligned}
E(\widehat{T}_e | F_1, K_0) &= c(F_1, K_0) - \sum_{j \in K_0} a_j E(y_j^t) \\
&> c(F_1, K_1) - \sum_{j \in K_0} a_j E(y_j^t) - \sum_{\substack{j \notin K_0 \\ j \in K_1}} a_j E(y_j^t) \\
&= c(F_1, K_1) - \sum_{j \in K_1} a_j E(y_j^t) \\
&= E(\widehat{T}_e | F_1, K_1),
\end{aligned}$$

as required. **QED**

References

- Abreu, D. (1988): “On the Theory of Infinitely Repeated Games with Discounting”, *Econometrica* 56, 383-396.
- Acemoglu, D. and J.A. Robinson (2000): “Why did the West Extend the Franchise? Democracy, Inequality and Growth in Historical Perspective”, *Quarterly Journal of Economics* 115, 1167–1199.
- Aidt, T.S., J. Dutta and E. Loukoianova (2006): “Democracy Comes to Europe: Franchise Extension and Fiscal Outcomes 1830–1938”, *European Economic Review* 50, 249–283.
- Alesina, A., and D. Rodrik (1994): “Distributive Politics and Economic Growth”, *Quarterly Journal of Economics* 109, 465–90.
- Alston, L.J. and J.P. Ferrie (1993): “Paternalism in Agricultural Labor Contracts in the U.S. South: Implications for the Growth of the Welfare State”, *American Economic Review* 83, 852–876.
- Atkinson, A.B. (1995): “The Welfare State and Economic Performance”, *National Tax Journal* (June), 171–198.
- Atkinson, A.B. (2004): “Income Tax and Top Incomes Over the Twentieth Century”, *Hacienda Publica Española/Revista de Economía Publica* 168, 123–141.
- Atkinson, A.B. (2007): “The Distribution of Top Incomes in the United Kingdom 1908–2000”, in A.B. Atkinson and T. Piketty (eds.), *Top Incomes Over the 20th Century*, Oxford: Oxford University Press.
- Atkinson, A.B. and A. Leigh (2010): “The Distribution of Top Incomes in Five Anglo-Saxon Countries over the Twentieth Century”, IZA DP No. 4937.
- Barr, N. (2004): *Economics of the Welfare State*, 4th ed., Oxford: Oxford University Press.
- Benabou, R. (1996): “Inequality and Growth”, *NBER Macroeconomics Annual*, Cambridge: MIT press.
- Benabou, R. (2000): “Unequal Societies: Income Distribution and the Social Contract.” *American Economic Review* 90, 96–129.
- Bjorklund, A., M. Palme and I. Svensson (1995): “Tax Reforms and Income Distribution: An Assessment Using Different Income Concepts”, *Swedish Economic Policy Review*, 229–266.
- Bridgen, P. (2006): “A Straightjacket with Wriggle Room: The Beveridge Report, the Treasury and the Exchequer’s Pension Liability, 1942-59”, *Twentieth Century British History* 17(1): 1–25.
- Budget of the United States: Historical Tables Fiscal Year 2005, Table 2.3.

- Cronin, J.E. (1979): “The Peculiar Pattern of British Strikes Since 1888”, *Journal of British Studies* 13, 118–141.
- Dal Bó, E., P. Dal Bó and J. Snyder (2009): “Political Dynasties”, *Review of Economic Studies* 76, 115–142.
- Domhoff, G.W. (1990): *The Power Elite and the State: How Policy is Made in America*, New York: Aldine de Gruyter.
- Domhoff, G.W. (1996): *State Autonomy or Class Dominance? Case Studies on Policy Making in America*, New York: Aldine de Gruyter.
- Flora, P. (1983): *State, Economy and Society in Western Europe 1815–1975: a Data Handbook*, Frankfurt: Campus Verlag.
- Gilens, M. (2005): “Inequality and Democratic Responsiveness”, *Public Opinion Quarterly* 69, 778–896.
- Heclo, H. (1974): *Modern Social Politics in Britain and Sweden*, New Haven: Yale University.
- Kerbo, H.R. and R.A. Shaffer (1986): “Unemployment and Protest in the United States, 1890-1940: A Methodological Critique and Research Note”, *Social Forces* 64, 1046–1056.
- Kollman, G. (1996): “Summary of the Major Changes in Social Security Cash Benefits Program: 1935-1996”. Congressional Research Service Report for Congress, <http://www.ssa.gov/history/pdf/crs9436.pdf>.
- Korpi, W. (1980): “Social Policy and Distributional Conflict in the Capitalist Democracies. A Preliminary Comparative Framework”, *West European Politics* 3, 296–316.
- Korpi, W. and J. Palme (2003): “New Politics and Class Politics in the Context of Austerity and Globalization: Welfare State Regress in 18 Countries, 1975-95”, *American Political Science Review* 97, 425-446.
- Kuznets, S. (1953): *Shares of Upper Income Groups in Income and Savings*, New York: NBER.
- Kuznets, S. (1963): “Quantitative Aspects of the Economic Growth of Nations: VIII. Distribution of Income by Size”, *Economic Development and Cultural Change* 11, 1–80.
- Lampman, R.J. (1984): *Social Welfare Spending*, Orlando: Academic Press.
- Lindert, P.H. (2004): *Growing Public*, Cambridge: Cambridge University Press.
- Lowe, R. (1999): *The Welfare State in Britain since 1945*, 2nd ed., New York: St. Martin’s.
- Lundberg, U. and K. Åmark (2001): “Social Rights and Social Security: the Swedish Welfare State, 1900-2000”, *Scandinavian Journal of History* 26, 157–176.
- Mares, I. (2003): *The Politics of Social Risk*, Cambridge: Cambridge University.

- Meltzer, A.H. and S.F. Richard (1981): “A Rational Theory of the Size of Government.” *Journal of Political Economy* 89, 914–927.
- OECD (2008): *Revenue Statistics 1965-2007*, Tax Levels and Tax Structures, 1965-2006.
- Perry, M. (2000): *Bread and Work: Social Policy and the Experience of Unemployment, 1918-1939*, London: Pluto Press.
- Persson, T., and G. Tabellini (1994): “Is Inequality Harmful for Growth?”, *American Economic Review* 84, 600–21.
- Piketty, T. and E. Saez (2003): “Income Inequality in the United States, 1913-1998”, *Quarterly Journal of Economics* CXVIII, 1–39.
- Piketty, T. and E. Saez (2007): “How Progressive is the U.S. Federal Tax System? A Historical and International Perspective”, *Journal of Economic Perspectives* 21, 3–24.
- Putnam, R. D. (1976): *The Comparative Study of Political Elites*, Prentice Hall.
- Quadagno, J.S. (1984): “Welfare Capitalism and the Social Security Act of 1935”, *American Sociological Review* 49, 632–647.
- Rauchway, E. (2008): *The Great Depression & the New Deal: A Very Short Introduction*, New York: Oxford University.
- Rodrik, D. (1998): “Why Do More Open Economies Have Bigger Governments?”, *Journal of Political Economy* 106, 997–1032.
- Romer, C.D. and D.H. Romer (2012): “The Incentive Effects of Marginal Tax Rates: Evidence from the Interwar Era”, NBER Working Paper No. 17860.
- Saez, E. (2004): “Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications”, in J.M. Poterba (ed.), *Tax Policy and the Economy*, vol. 18, 117–173.
- Statistics Sweden (2008): *Distribution of Income 1975–2007*, Household Finances Statistics.
- Steinmo, S. (1993): *Taxation and Democracy: Swedish, British and American Approaches to Financing the Modern State*, New Haven: Yale University Press.
- Thane, P. (1984): “The Working Class and State ‘Welfare’ in Britain, 1880-1914”, *The Historical Journal* 27, 877–900.
- Twight, C. (1995): “Evolution of Federal Income Tax Withholding: The Machinery of Institutional Change”, *The Cato Journal* 14, 359–395.
- U.S. Senate (July-August 1942) Comm. on Finance, Revenue Act of 1942. Hearings [On H.R. 7378], Vol. 1, 77–2.
- Valocchi, S. (1989): “The Relative Autonomy of the State and the Origins of British Welfare Policy”, *Sociological Forum* 4, 349-365.

Weaver, C.L. (1983): “On the Lack of a Political Market for Compulsory Old-Age Insurance Prior to the Great Depression: Insights from Economic Theories of Government”, *Explorations in Economic History* 20, 294–328.

Fig. 1. *Social Transfers*

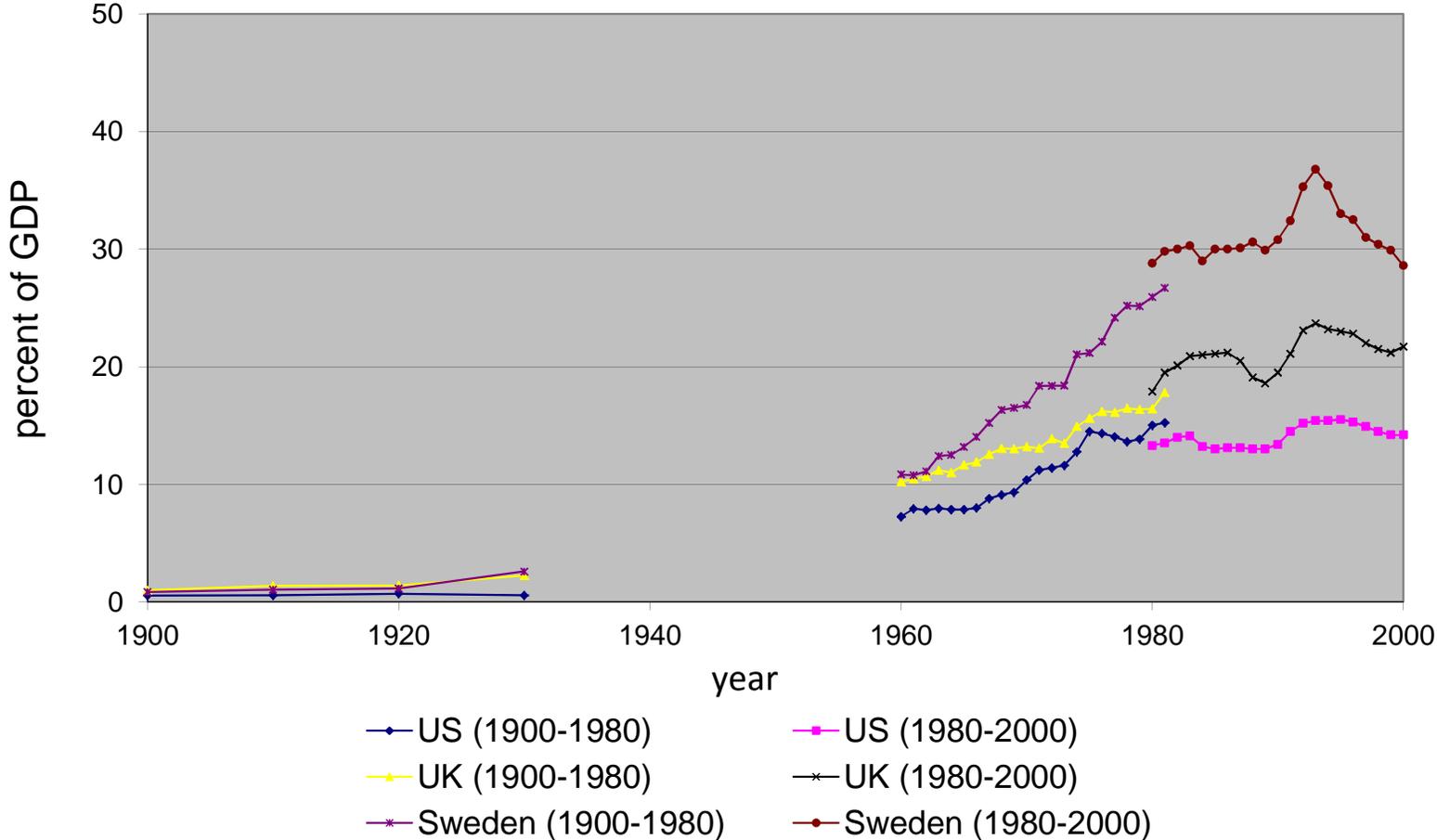


Fig. 2. *Top 1% Effective Tax Rates*

