Economic Coercion in Public Finance

Beat Hintermann, University of Basel

Thomas F. Rutherford, University of Wisconsin at Madison

June 20, 2013

Draft document – please do not cite

Abstract

We develop a theory of social planning with a constraint on economic coercion, which we define as the difference between consumers' actual utility, and the "counterfactual" utility they expect to obtain if they were able to set policy themselves. The social planner limits economic coercion, either to protect minorities or to prevent disenfranchised groups from engaging in socially costly behavior. We show that if consumers are not fully rational and/or informed, the introduction of a coercion constraint leads to additional terms that render counterfactual utility endogenous. The coercion-constrained policy optimum appears to be dominated by a set of policies that increase overall welfare as well as the utility of the most-coerced individuals, but carrying out such a policy change would in fact change the counterfactual and increase coercion beyond the permissible level. We obtain similar results in a probabilistic voting framework, but their interpretation differs: Whereas the social planner will use his superior knowledge about the effect of policy variables to increase long-term welfare subject to a coercion maximum, the political candidate will employ the same knowledge to win elections, possibly to the long-term disadvantage of voters. Economic coercion can also lead to a divergence between announced and realized policies. We illustrate our results numerically.

1 Introduction

There are many forms of coercion,[1] including physical, psychological, legal, religious, sexual etc. In an economic context, coercion occurs wherever individuals are subject to decisions which directly affect their utility, but over which they have no control. Citizens of a nation-state are coerced into accepting, for example, a particular level of a public good or a tax schedule that will generally differ from what they would prefer.[2]

Even though members of a society are coerced by group rules, joining the group may convey greater utility for an individual than remaining outside, mainly due to coercion of others [Baumol 2003]. Subjection to economic coercion by joining a group can therefore be entirely voluntary (being member of the group while not being subjected to its laws would of course be even better from an individual's perspective, but this is assumed to be infeasible in a non-dictatorial society). In this paper we focus on the situation where a group has already been formed; for a treatment of economic coercion in the context of coalition formation, see e.g. Gamson [1961] and Skarpedas [1992].

The economic subfield where coercion within an existing group is acknowledged most explicitly is the normative theory of public choice, which is concerned with the definition of optimal decision rules for policy. This literature goes back to the work of Wicksell [1896] and Lindahl [1919], who proposed approximate unanimity as a direct consequence to the desire of minimizing coercion. Buchanan and Tullock [1962] derive unanimity in the framing of a constitution as an efficiency condition, but without providing a formal definition of economic coercion.

Our paper is based on the framework developed by S. Winer, G. Tridimas and W. Hettich [2008], who to the best of our knowledge are the first to introduce an explicit concern with coercion into social planning. We adopt their working definition of coercion as the difference between the hypothetical utility level a consumer/voter would achieve if she were in control of some (or all) policy variables, and the utility level she actually obtains conditional on actual policy. In keeping with their language, we refer to this hypothetical utility level as a consumer's "counterfactual" utility. From a social planning

[1] Coercion is the "act, process or power of coercing", with the definition of the relevant verb given by Merriam-Webster as 1.) to restrain or dominate by force, 2.) to compel to an act or choice, or 3.) to achieve by force or threat.

[2] Economic coercion is often discussed in the context of trade sanctions between sovereign nation-states or in a framework of international trade. In this paper, we focus on economic coercion within societies, rather than between them.

point of view, economic coercion is the price we pay to reach a social optimum, given heterogeneous preferences and endowments. It can therefore not be the goal to eliminate coercion altogether, but there may be reasons to set limits to the amount of coercion any particular individual or group may be subjected to.

We develop our model both in a social planning and a probabilistic voting framework. A concern with coercion affects the outcome in both contexts, but the results and their interpretation differ in important aspects. The benevolent social planner may want to limit the extent to which policy choices alienate segments of the population in order to prevent them from organizing labor strikes or mass protests, or engaging in illegal activities such as refusal to pay taxes, vandalism or terrorism. Even though social welfare will be reduced by imposing a binding coercion constraint, the reduction in collateral damages may more than make up for the welfare loss. In contrast, a political candidate's motivation to limit coercion is simply the desire to get elected, and he will therefore limit coercion where this transforms into the highest vote gain.

We show that under full rationality and complete information, a concern with coercion can be addressed within the social welfare function by adding additional welfare weight to the most coerced groups. If, however, consumers make mistakes when computing their counterfactual, the latter becomes endogenous, and the social planner has to consider the change in counterfactual in response to a change in policy.

Incorporating coercion into social planning can be interpreted as a kind of paternalism related to consumers' counterfactual utility, which we label *counterfactual paternalism*: At the coercion-constrained policy optimum, there will generally exist alternative policies that would increase overall welfare and which the most coerced consumers think they prefer, based on a counterfactual computed at the current policy. However, carrying out this policy change would lead these consumers to adjust their counterfactual in a way that increases their level of economic coercion beyond the coercion constraint. If counterfactual errors are sufficiently large, and the coercion constraint sufficiently tight, the policy outcome may even be allocatively inefficient in the sense that it is inconsistent with social welfare maximization based on a set of nonnegative welfare weights. We derive a necessary condition for allocational inefficiency, but we are not able to identify a set of sufficient conditions that is generally valid.

In the probabilistic voting context, a political candidate uses his superior knowledge about counterfactuals differently: He will propose a platform that caters to voters' direct utilities without taking into account the effect of the policies on coercion. He has an incentive to do so because coercion will change only after the election, when policy is put into place. The candidate will give voters exactly what they think they want, even if this is not in their long-term interest. We refer to this outcome as *counterfactual populism*.

Once elected, the politician has to choose a policy. We show that a concern with economic coercion will generally lead him to choose a policy level that differs from platform that he previously ran on, because the actual policy (in contrast to a pre-election policy platform) will change counterfactuals, which affects his re-election probabilities.

Our results are linked to the emerging literature of behavioral public economics, which aims to incorporate systematic deviations between observed behavior on the one hand, and behavior that is consistent with neoclassical principles such as utility maximization by fully informed and rational, forward-looking agents.[3] We argue that consumers/voters compare their actual situation with what they would like the world to be like. The anomaly that we postulate is that consumers make systematic mistakes when computing counterfactual utility, and develop a framework to incorporate this situation into decision making by a benevolent government or a self-interested political candidate.

In the next section, we develop a model of economic coercion in a social planning context. In section 3, we re-cast the model in a probabilistic voting framework. Section 4 contains our application, and section 5 concludes.

2 Coercion in social planning

2.1 Model

We describe consumers' preferences with the utility function

$$U^h(X_{hi}; G) \qquad\qquad h = 1, ..., H; \quad i = 1, ..., N$$

where X_{hi} refers to demand for (supply of) the good (factor) i by a consumer of type h, and G is a public good. [4] We denote aggregate quantities by $X_i \equiv \sum_H X_{hi}$.

[3] For a review of this literature, see Bernheim and Rangel [2007].

[4] We refer to consumer types rather than individual consumers in order to facilitate comparison with the probabilistic voting model in section 3 (the types become interest groups), and the numerical application in section 4. Furthermore, G could be a vector and include environmental quality and other measures that affect

Consumers treat the level of public provision as given and maximize utility by solving

$$\max_{X_{hi}} U^h(X_{hi}; G) \quad s.t. \quad \sum_{i=1}^{N} P_i X_{hi} \leq I_h \tag{1}$$

where I_h is a fixed amount of (non-taxable) lump-sum income, and P_i refers to the consumer price, which may include a tax. Solving (1) gives rise to the indirect utility function

$$V^h(P_i, I_h, G) = U^h[X_{hi}(P_i, I_h; G); G]$$

If production is efficient,[5] we can represent technology by means of the aggregate production function $F(X_i; G)$. Substituting market clearance into the production function yields an implicit production-possibilities frontier:

$$F(X_i; G) = F\left(\sum_{h=1}^{H} X_{hi}(\cdot); G\right)$$

The government funds production of the public good by means of ad-valorem taxes on the private goods and factors, with good 1 serving as the numeraire:

$$P_i = p_i(1 + t_i) \ \text{ for } \ i = 2, ..., N$$
$$t_1 = 0; \quad P_1 = p_1 = 1$$

With efficient production, net-of-tax prices p_i are equal to the MRT between these goods (factors) and the numeraire. Tax revenue is fully used to finance the public good at a constant price p_G:

$$p_G \cdot G = \sum_{i=2}^{N} p_i \, t_i \cdot X_i \tag{2}$$

Using this identity, the social planner maximizes social welfare by choosing a vector of tax rates t subject to the technology constraint and a constraint on the maximally allowable amount of economic coercion that a particular person (or group) may be subjected to. We use the difference between actual welfare V^h and counterfactual welfare $\tilde{V}^h$ as

utility, and over which the consumer has no control. In our numerical application, we focus on the case of one pure public good and a dirty intermediate good that produces an aggregate environmental externality.

[5] Firms equalize the marginal rate of technical substitution between factors, the marginal rate of transformation between goods and the marginal product of any factor in the production of all goods.

our definition of economic coercion:[6]

$$\max_{t} \quad W = \sum_{h=1}^{H} n_h \cdot \alpha_h \cdot V^h(P_i; I_h; G(\cdot)) \tag{3}$$

$$s.t. \quad F\bigg(\sum_{h=1}^{H} X_{hi}(\cdot); G(\cdot)\bigg) \leq 0$$

$$\tilde{V}^h - V^h \leq \bar{K}_h \quad \forall\ h$$

The social planner places weight α_h on consumer group h, with $\sum_{j=1}^{H} \alpha_j = 1$ and n_h referring to the number of consumers in each group. $\bar{K}_h$ is the upper limit of economic coercion to which goupe h may be subjected, and which is determined in the political process along with the welfare weights.[7] The Lagrangian and first-order necessary condition w.r.t. to element $t_i \in t$ are

$$L = \sum_{h=1}^{H} n_h \alpha_h V^h - \lambda F(\cdot) + \sum_{h=1}^{H} n_h \kappa_h \big[\bar{K}_h - (\tilde{V}^h - V^h)\big] \tag{4}$$

$$\sum_{h=1}^{H} n_h \bigg[(\alpha_h + \kappa_h) \cdot \frac{\partial V^h}{\partial t_i} - \kappa_h \frac{\partial \tilde{V}^h}{\partial t_i}\bigg] = \lambda \cdot \frac{\partial F(\cdot)}{\partial t_i} \tag{5}$$

Before further analyzing (5), we turn to the counterfactual, $\tilde{V}^h$. The Lagrangian of the counterfactual utility maximization problem is

$$\max_{X_{hi}; t^h} \quad \tilde{L}^h = U^h\big(X_{hi}; G(\cdot)\big) - \tilde{\mu}_h\bigg(\sum_{i=1}^{N} P_i X_{hi} - I_h\bigg) - \tilde{\lambda}_h \tilde{F}(\cdot) \tag{6}$$

where $\tilde{t}^h$ is a vector of policy variables that the consumer "controls" in the counterfactual problem, and the term multiplied by $\tilde{\mu}_h$ is the consumers' budget constraint. Solving (6) and substituting the derived demands into the utility function yields the indirect counterfactual utility

$$\tilde{V}^h = \tilde{V}^h[I_h; \epsilon_h(X_i; t)] \tag{7}$$

[6] Equivalently, we could transform utility by $U^* = e^U$ and use a proportional coercion constraint.

[7] On normative grounds, it is not clear why the coercion limit should differ across consumer groups and could therefore be replaced by $\bar{K}_h = \bar{K}\ \forall\ h$. However, in a political economy context this may be different. To keep the model general, we allow the coercion limit to vary across consumer types. Note also that the coercion constraint will generally be binding for one group only, such that there is no qualitative difference between a uniform and individual coercion constraints.

where ϵ_h refers to errors that consumers may make when solving (6), which may depend on observed demand levels and prices, and thus on the government's choice of t. The presence or absence of such an error turns out to be crucial for limiting economic coercion.

Consumers may or may not use the same vector of control variables as the social planner, which is the first source of error in consumers' counterfactual. The "tilde" over the production function $\tilde{F}(\cdot)$ implies that consumers may consider a different technology, market clearance and budget constraint as the government, adding a second type of potential counterfactual error. Even if consumers apply the same set of policy variables and constraints as the government, they may still compute different first-order conditions if they use approximations. This is a third source of error, and the one that we will focus on in our application.

Now we return to the social planner's problem. Application of the envelope theorem implies that the derivative of the counterfactual utility $\tilde{V}^h$ w.r.t. a policy variable in t is given by the corresponding derivative of the counterfactual Lagrangian, evaluated at the counterfactual demand level and policy choice computed by consumers:

$$\frac{\partial \tilde{V}^h}{\partial t_i} = \left.\frac{\partial \tilde{L}^h}{\partial t_i}\right|_{\tilde{X}^*_{hi}, t^{h*}} \tag{8}$$

If consumers solve the counterfactual problem correctly, it must be that $\partial \tilde{L}^h / \partial t_i = \partial \tilde{L}^h / \partial t_i^h = 0 \; \forall \; i$, and all terms associated with consumers' counterfactual utility drop out of the social planner's first-order conditions (5). If, however, consumers make mistakes when computing the counterfactual and these mistakes are sensitive to marginal policy choices, (8) is not zero, and the social planner needs to take the change in counterfactual utility w.r.t. a change in a policy variable into account when maximizing social welfare.

2.2 The effect of coercion on the equilibrium outcome

Condition (5) shows that if consumers make no mistakes, the welfare weight $\alpha_h \geq 0$ for type h is simply increased to $\alpha_h + \kappa_h$.[8] In other words, a concern with coercion places additional welfare weight on the consumer type for which the coercion constraint is binding (otherwise, $\kappa_h = 0$), but all terms related to counterfactual utility drop out of the model. Intuitively, correctly computed counterfactuals are a function only of primitives

[8] The coercion constraint will be binding for more than one consumer type only by coincidence.

like consumer preferences, technology and availability of resources. Since these primitives are fixed, counterfactual utility becomes a constant, and like any constant drops out of the optimization problem.

The same is true if consumers compute wrong counterfactuals which are fixed, or counterfactuals that may be wrong globally but share the same local first-order conditions. In both cases, $\partial\tilde{V}^h/\partial t_i = 0 \ \forall \ i$, and the effect of coercion is simply to place additional welfare weight on the coerced consumer type.

However, consumer mistakes or approximations that make the counterfactuals a function of policy variables lead to additional terms in the first-order conditions, which can be interpreted as welfare weights that differ across policy dimensions. These weights can be positive or negative, depending on the coercion response to a change in a policy variable. We can now state our first result:

Proposition 1. *a.) If consumers correctly compute their counterfactual utility, introducing a coercion constraint adds positive welfare weight to consumers for whom the coercion constraint is binding. These welfare weights are constant across policy dimensions. The same is true for incorrectly computed counterfactuals that are fixed either globally or locally, such that $\partial\tilde{V}^h/\partial t_i = 0 \ \forall \ i$.*

b.) If consumers make mistakes when computing their counterfactual utility and these mistakes depend on marginal changes in the social planner's choice vector, introducing a coercion constraint introduces additional welfare weights that differ across policy dimensions.

c.) Consumer type h receives a negative *welfare weight along policy dimension $t_i \in t$ if*

$$\alpha_h < \kappa_h \cdot \left(\frac{\partial\tilde{V}^h/\partial t_i}{\partial V^h/\partial t_i} - 1 \right) \tag{9}$$

d.) Consumer type h will be worse off due to the introduction of a binding coercion constraint if

$$\kappa_h \leq \alpha_h \cdot \sum_{j=1}^{H} \kappa_j \tag{10}$$

In particular, this is true for all consumer types for which $\kappa_h = 0$.

Proof. Parts a.) and b.) follow from the discussion in the text. Part c.) is based on rewriting (5) and setting the terms in brackets equal to zero. Part d.): Consumer h's overall welfare weight is α_h in the absence of a concern with coercion, and $(\alpha_h + \kappa_h)/(1 + \sum \kappa_h)$ with a binding coercion constraint (recall that $\sum_{j=1}^{H} \alpha_j = 1$). The latter is less than or equal to the former if eq. (10) holds. Since $\sum_h \partial W/\partial \bar{K}_h = \sum_h n_h \kappa_h > 0$ for a binding constraint, total welfare decreases when placing a limit on economic coercion, and type h receives a smaller (or at most an equal) share of it as before. □

To provide some intuition about Proposition 1, note that condition(9) can only occur if a change in t_i affects actual and counterfactual utility in the same direction, and the marginal effect on the latter is larger than on the former. Suppose, for example, that $\partial V^h/\partial t_i > 0$ at the equilibrium, because the utility gain through an increase in the public good more than offsets the utility loss from a higher price, net of all general equilibrium effects. If counterfactual utility, and thus coercion, increases by more than actual utility and α_h/κ_h is sufficiently low such that (9) holds, the government will consider this consumer type's preferences negatively when computing the optimal tax on X_i. The reason is that the government knows that the increase in welfare due to an increase in the consumer's actual welfare is more than offset by the increase in economic coercion.

2.3 Implications for welfare and for efficiency

By construction, adding a binding coercion constraint cannot increase welfare. However, the extent of the welfare loss depends not only on the stringency of the coercion constraint, but also on the nature of consumers' counterfactuals. If the counterfactual error is large enough, the solution may even be inefficient in purely allocative terms, i.e. irrespective of any particular social welfare function.

Figure 1 illustrates, using a simplified setting where the government controls two policy instruments (this is the context our numerical model in Section 4: A labor tax t_L and a tax t_E on an externality-generating good, such that every point in the figure reflects a specific policy combination. There are three consumer types h, i and j who differ in their tastes and endowments, and their utility maximum is represented by the corners of the triangular shape.

The borders of this shape are the contract curves between any two types, which are found by placing zero welfare weight (and a nonbinding coercion constraint) on the third type and varying the welfare weights between zero and one.[9] We call the area enclosed by

[9] More precisely, the contract curve is the locus of points where the indifference curves of the two types are

the contract curves the *Pareto-optimal policy space*, because any solution within this space is allocationally efficient and could be generated as an outcome involving a particular set of nonnegative welfare weights in the absence of a coercion constraint.

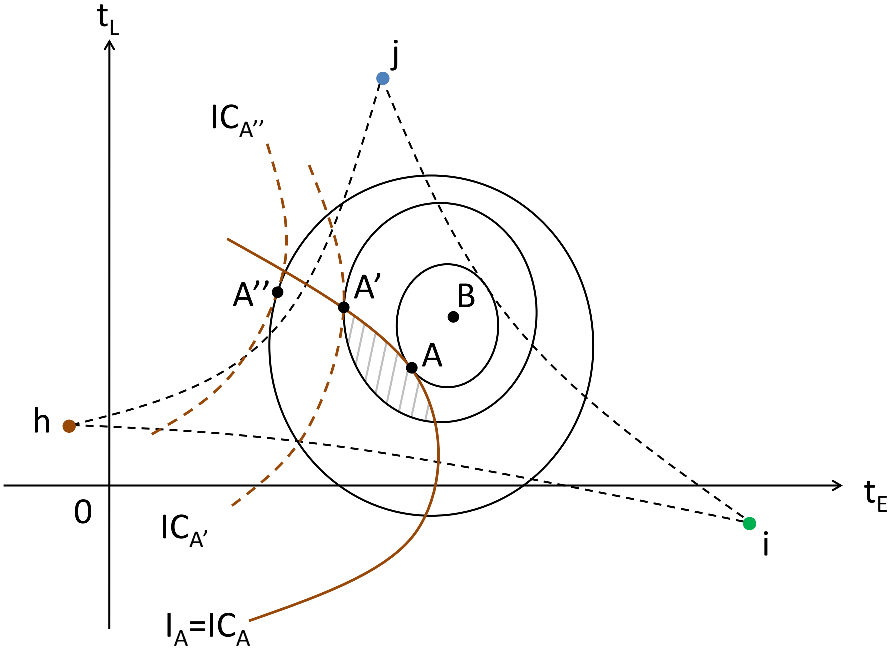

Figure 1: Coercion-constrained policy outcome

Point B represents the coercion-unconstrained welfare maximum, and the lines around it are social indifference curves. Suppose the coercion constraint is binding for type h, but not for types i and k. With a correct counterfactual, h's counterfactual utility is located at her true optimum (the dot in the figure). Since $\partial\tilde{V}^h/\partial t = 0$, the counterfactual is independent of actual policy, making type h's indifference curves also coincide with her iso-coercion curves. The coercion-constrained policy outcome A is located at the tangency between the social indifference curves and type h's iso-coercion curve IC$_A$ for a coercion level of $V^h_{opt} - V^h = \bar{K}_h$. The welfare loss from introducing the coercion constraint is the difference between welfare at the social optimum and the coercion-constrained outcome, W(B)-W(A).

If consumers do make mistakes when computing their counterfactual and the latter depends on current policy such that $\partial\tilde{V}^h/\partial t \neq 0$, the iso-coercion curves will no longer be tangent to the iso-utility curves. Even a marginal move from any equilibrium changes the counterfactual, which implies that moving along an iso-utility curve necessarily changes

tangent.

the level of economic coercion.[10]

The coercion-constrained solution is again at the tangency between the social indifference curve and type h's relevant iso-coercion curve, for example $IC_{A'}$, leading to a solution at A'. The resulting welfare loss in this example is W(B)-W(A').

The endogeneity of the counterfactual leads to a welfare loss in itself. To show this, suppose we could somehow fix the counterfactual associated with A', such that $\partial\tilde{V}^h/\partial t = 0$ and the iso-coercion curve coincides with type h's indifference curve I_A. The resulting coercion-constrained policy outcome would be A, with W(A)>W(A') due to an increase in aggregate utility for types i and j (since type h's utility remains constant along I_A). The welfare loss due to the endogenous nature of the counterfactual is therefore W(A)-W(A'). Intuitively, the move from A' to A cannot take place because it (erroneously) increases counterfactual utility, resulting in coercion exceeding the constraint $\bar{K}$.

More generally, welfare is decreased if the iso-coercion curve is not tangent to the respective type's indifference curve at the coercion-constrained policy outcome, relative to a situation where the two curves have the same slope. Tangency between iso-coercion and iso-utility curves can arise due to three reasons: 1.) correct counterfactuals, 2.) incorrect but fixed counterfactuals, and 3.) globally incorrect but locally correct counterfactuals.[11] In each of these cases, $\partial\tilde{V}/\partial dt_i = 0 \;\forall\; i$.

If consumer errors are large and the coercion constraint sufficiently tight, an errouneous counterfactual can lead to a policy equilibrium that is not only distributionally, but even allocationally inefficient. An example of such an outcome is point A" in Figure 1, located at the tangency between the iso-coercion curve $IC_{A''}$ and the corresponding social indifference curve. Welfare is decreased by W(B)-W(A"), but unlike the equilibria at A and A', the solution at A" could not be replicated by a set of nonnegative welfare weights, since all such solutions have to lie within the Pareto-optimal policy space. An equilibrium at a point like A" corresponds to an overall negative welfare weight for at least one agent, which can happen if (and only if) condition (9) applies to at least one policy dimension for at least one consumer type.[12]

[10] Formally, the slope of the iso-utility curve can be derived by setting $dV = \partial V/\partial t_L dt_L + \partial V/\partial t_E dt_E = 0$ and solving for $\frac{dt_L}{dt_E} = -\frac{\partial V/\partial t_E}{\partial V/\partial t_L}$. The slope of the iso-coercion curve is $\frac{dt_L}{dt_E} = -\frac{\partial\tilde{V}/\partial t_E - \partial V/\partial t_E}{\partial\tilde{V}/\partial t_L - \partial V/\partial t_L}$. The two slopes only coincide if $\partial\tilde{V}/\partial t_L = \partial\tilde{V}/\partial t_E = 0$.

[11] A Taylor series approximation of counterfactual to actual utility would have this property. Whereas the global counterfactual could be quite far away from the actual optimum, the (negative) change in counterfactual utility coincides with the change in actual utility for infinitesimally small changes in the policy instruments.

[12] So far we have not been able to establish a generally applicable sufficient condition, because the overall

Figure 2 shows the region around the inefficient equilibrium A". Moving inside the shaded area would not only increase social welfare and type h's utility, but also efficiency. However, this move cannot take place due to h's erroneous counterfactual: A move to the lower right increases counterfactual utility by more than actual utility, thus increasing economic coercion beyond the level associated with IC$_{A''}$.

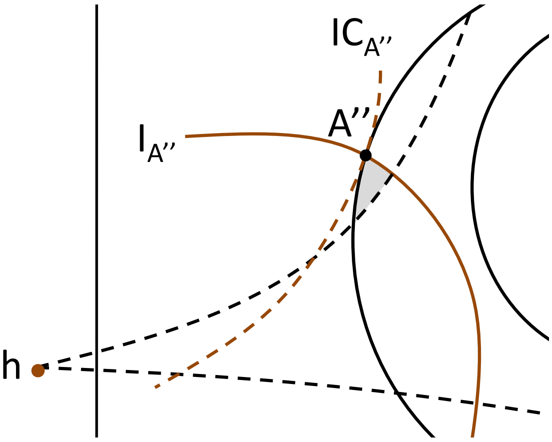

Figure 2: Allocationally inefficient outcome

The coercion-constrained outcome under an erroneous counterfactual can be interpreted as a type of paternalism, but one that refers to consumers' counterfactual rather than actual utility function. We refer to this situation as *counterfactual paternalism*: Not taking the changing nature of the counterfactual into account, consumer type h thinks she would prefer a move from A' inside the shaded area in Figure 1, or from A" inside the shaded area in Figure 2. The social planner does not allow such a move even though it would increase overall welfare (and in case of Figure 2 even efficiency), because he knows that this would result in increased economic coercion for type h, once the counterfactual adjusts.

2.4 Minimizing coercion

Unless consumers have homogenous preferences and identical endowments, every policy choice will lead to some coercion since policy variables apply to everybody. A certain level of coercion is therefore unavoidable, even if the government single-mindedly pursues a policy of minimizing coercion. The minimum level of coercion for consumer type h can

welfare weight that a consumer type receives depends not only on the partial derivatives of actual and counterfactual utility w.r.t. all policy variables at the equilibrium, but also on the relative importance each policy variable has on utility in absolute terms. For example, even if condition (9) holds strongly along policy dimension t_j for a consumer type, this will not lead to an overall welfare weight if this policy variable is not very important in determining her utility.

be found be minimizing $\bar{K}_h$ subject to the production possibility frontier and the coercion constraint:

$$\min_{t} \quad \bar{K}_h \qquad s.t. \quad F(\cdot) \leq 0 \\ \tilde{V}^h - V^h \leq \bar{K}_h \tag{11}$$

If we denote social welfare associated with the solution of problem (11) as $\bar{W}$, and welfare without a concern for coercion as W^0, the maximum decrease in social welfare due to a concern with coercion for group h is given by $W^0 - \bar{W}$.

Figure 3 shows the relationship between the coercion constraint $\bar{K}_h$ and social welfare. The slope of this figure is given by differentiating 4 w.r.t. the coercion constraint:

$$\frac{\partial W}{\partial \bar{K}_h} = \frac{\partial L}{\partial \bar{K}_h}\bigg|_{t^*} = \kappa_h$$

The minimum attainable level of coercion is given by $\bar{K}_h^{min}$, which leads to a level of social welfare of $\bar{W}$. Coercion-unconstrained social welfare is maximized at the point $\left(\bar{K}_h^0, W^0\right)$, where all shadow prices are zero. Increasing $\bar{K}_h$ beyond this point has no effect on welfare in our model (solid line), since we specified the constraint on coercion as a weak inequality. Note that if we instead had introduced a constraint of an exact equality, then increasing $\bar{K}_h$ beyond $\bar{K}_h^0$ would decrease social welfare (dashed line).

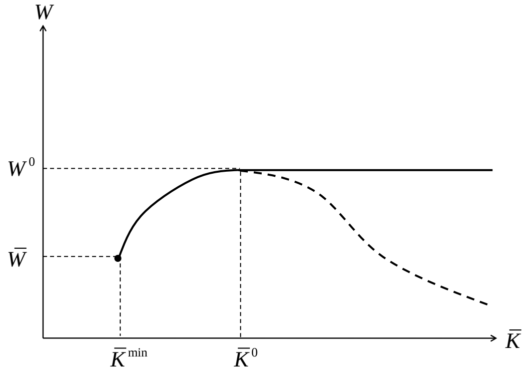

Figure 3: Social welfare as a function of the constraint on coercion

3 Coercion in a probabilistic voting framework

The model in the previous subsection can be re-cast in a probabilistic voting context of two-party competition with interest groups, following Mueller [2003, Ch. 12]. The structure of the outcome remains essentially the same, but the interpretation differs, as well as the motivation for introducing a concern with coercion in the first place.

3.1 Implicit welfare maximization with probabilistic voting

Assume that preferences and endowments are homogeneous within, but not across interest groups $h = 1, ..., H$. Voters have preferences for one of the two candidates/parties, called a political bias. The bias in favor of candidate 1 by voter j belonging to interest group h is given by b_{hj} and drawn from a probability function specific to interest group h. A positive b_{hj} represents a bias in favor of candidate 1, and $b_{hj} < 0$ is a bias in favor of candidate 2.

Let V_1^h and V_2^h represent the utility that a voter in interest group h receives from the policy platform of candidate 1 and 2, respectively, and π_{1hj} is the probability (from the candidates' point of view) that a particular voter j in interest group h votes for candidate 1. Because there are only two parties and we assume that everybody votes, $1 - \pi_{1hj}$ is the probability that individual j votes for candidate 2. In the absence of a concern with economic coercion, voting (from voters' point of view) is determined by

$$\begin{aligned} \pi_{1hj} &= 1 \qquad \text{if} \quad E[V_1^h] > E[V_2^h] - b_{hj} \\ &= 0 \qquad \text{if} \quad E[V_1^h] \leq E[V_2^h] - b_{hj} \end{aligned}$$

The expectation terms reflect the fact that voters make a decision based on expected utilities after the election, when the promised policies are in place. Defining $Z_h \equiv E[V_1^h] - E[V_2^h]$ as the expected utility differential from the two platforms for interest group h, the probability that voter j votes for candidate 1 can be written as

$$\begin{aligned} E[\pi_{1hj}] = E[\pi_{1h}] &= \text{Pr}\ (Z_h > -b_{hj}) = \text{Pr}\ (b_{hj} > -Z_h) \\ &= 1 - \text{Pr}\ (b_{hj} \leq -Z_h) = 1 - f_h(-Z_h) \end{aligned}$$

where $f_h(\cdot)$ refers to the cumulative density function (cdf) of b_h. Candidate 1 chooses the policy vector t_1 to maximize his expected vote share VS_1 over all interest groups with membership n_h subject to the same (implicit) production-possibility constraint as the social planner in the previous section, where the government budget constraint (2) has been substituted in:

$$\max_{t_1} E[VS_1] = \sum_{h=1}^{H} n_h \cdot E[\pi_{1h}] \quad s.t. \quad F(\cdot) \leq 0 \tag{12}$$

Because $\pi_{1h} = 1 - \pi_{2h}$, candidate 2's optimization problem is to choose t_2 to minimize (12). In equilibrium, both candidates will propose the same policy platform. The FONC w.r.t. policy dimension t_{1i} (candidate 1's choice of policy variable i) is

$$\sum_{h=1}^{N} n_h \cdot f'_h(-Z_h) \cdot \frac{dV_1^h}{dt_{1i}} = \lambda \frac{dF(\cdot)}{dt_{1i}} \tag{13}$$

where $f'_h(-Z_h)$ refers to the probability density function (p.d.f.) of b_h evaluated at $-Z_h$. This condition is the same that must be satisfied when maximizing a social welfare function of the form

$$W = \sum_{h=1}^{H} n_h \cdot f'_h\big(-Z_i|_{t=t_1^*}\big) \cdot V^h \tag{14}$$

subject to the same production-possibilities constraint, which establishes the result that competition of two political candidates in results in the implicit maximization of a social welfare function, which is an attractive normative property of this framework.

Note that Z_h has to be evaluated at the optimum from solving (13) defined by t_1^*, such that it becomes a constant; otherwise, maximization of (14) would lead to additional terms, and the two sets of FONCs would no longer be equivalent. In many applications, f_h is chosen to be the c.d.f. of the uniform distribution, in which case $f'_h = 1/(u_h - l_h)$ is a constant. The terms in the denominator refer to the upper resp. lower limits of the support of b_{hj}.

An important difference between social planning and probabilistic voting are the interpretation of the social welfare weights: Whereas the social planner either weighs utility of different groups equally (in the case of a Benthamite SWF) or assigns them welfare weights that are redistributive in nature, the implicit welfare weights in (14) depend on the density of the distribution of the bias.[13] Only by chance will there be a systematic re-

[13] For the uniform distribution, this is simply the inverse of the support: The more certain an interest group's voting behavior (i.e. the smaller the distance between l_h and u_h), the larger is f'_h and thus the greater its implied welfare weight, provided that $l_h \leq -Z_h \leq u_h$.

lationship between income levels and f'_h. The implicit welfare weight is higher for interest groups whose members are more responsive to the candidate's platform. In the political context, these are often referred to as swing voters. In contrast, voters that due to their political bias will likely vote for one of the two candidates do not figure prominently in the candidates' objective function. If candidate 1 is certain about the vote of a particular interest group (if $-Z_h$ is outside the support of x_{hj}), this interest group will receive an implicit welfare weight of zero, regardless of which candidate receives the vote. This is obviously quite different to the government-as-agent perspective. The two models coincide in their mathematical structure, but they generally lead to very different outcomes.

3.2 Coercion in probabilistic voting

Introducing a concern with coercion into the probabilistic voting model requires adding a second stochastic bias, one that captures the assumption that voters' disutility from a given level of economic coercion depends on who is in power.[14] We postulate that besides the expected utility differential Z_h, the voting decision by voter j in interest group h also depends on some function $Y_{khj}(\cdot), k = 1, 2$ of counterfactual utility $\tilde{V}_h$.

$$\begin{aligned} \pi_{1hj} &= 1 \qquad \text{if} \quad V_1^h + Y_{1hj}\big(V_1^h, E[\tilde{V^h}]\big) > V_2^h + Y_{2hj}\big(V_2^h, E[\tilde{V^h}]\big) - b_{hj} \qquad (15) \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

where we placed counterfactual utility in expectation signs to reflect the possibility that consumers make mistakes. Again, all terms are technically expectations of future utility taken at the time of voting, but since there is no uncertainty in the model, expectations are equal to realizations for actual utility.

To remain consistent with our approach in the social planning model, we use a coercion function that depends on the difference between (expected) counterfactual and actual utilities:

$$Y_{khj} = -\beta_{khj}\big(E[\tilde{V^h}] - E[V_k^h]\big); \qquad k = 1, 2 \qquad (16)$$

$$\beta_{hj} \equiv \beta_{2hj} - \beta_{1hj} = \beta_h \cdot b_{hj}; \quad \beta_h \geq 0$$

[14] For example, we believe that Tea-Party members in the USA may object to a given level of taxation differently depending on which party holds the presidency. Similarly, it may matter to Greek and Italian workers which party is in power when it comes to accepting austerity measures that are imposed externally.

A coercion bias in favor of party 1 is implied by $\beta_{hj} > 0$, and vice versa. Restricting the coercion bias to be proportional to the general bias in (16) ensures that they have the same sign.[15] If we drop the coercion-specific bias then $Y_{1hj} = Y_{1h} = Y_{2h}$, and coercion cancels out in (15). Transferring all stochastic terms to the RHS:

$$Z_h > \beta_{1hj}\big(E[\tilde{V}^h] - E[V_1^h]\big) - \beta_{2ih}\big(E[\tilde{V}^h] - E[V_2^h]\big) - b_{hj} \equiv -x_{hj} \tag{17}$$

The candidate's expected probability of receiving the vote from person j in interest group h is

$$E[\pi_{1h}] = Pr(Z_h > -x_{hj}) = 1 - g_h(-Z_h; V_1^h, V_2^h, \tilde{V}^h) \tag{18}$$

where $g_h(\cdot)$ is the c.d.f. of x_{hj} evaluated at $-Z_h$, but which depends on actual and counterfactual utilities.[16] The FONC from candidate's vote share maximization problem w.r.t. policy dimension t_{1i} are

$$\sum_{h=1}^{H} n_h\left(g_h' - \frac{\partial g_h}{\partial E[V_1^h]}\right) \cdot \frac{\partial E[V_1^h]}{\partial t_{1i}} - n_h \frac{\partial g_h}{\partial E[\tilde{V}^h]} \cdot \frac{\partial E[\tilde{V}^h]}{\partial t_{1i}} = \lambda \frac{\partial F(\cdot)}{\partial t_{1i}} \tag{19}$$

where g_h' is the p.d.f. of x_{hj} evaluated at the optimum.

Although voters may make mistakes when computing the counterfactual, they don't *expect* to make mistakes. This means that their expected counterfactual is not a function of t_1, even if their true counterfactual may well change once the policy t_1 is instituted. The last term on the LHS of (19) therefore drops out in the candidate's vote share maximization problem.

Before we proceed by comparing this to the associated implicit welfare maximization problem we need to understand the derivative of g_h w.r.t. actual and counterfactual utilities. In equilibrium, both candidates will choose the same policy platform such that $t_1^* = t_2^*$ and $V_1^h = V_2^h = V^h$. Substituting this into x_{hj} in (17) and using (16) leads to

$$\begin{aligned} x_{hj} &= b_{hj} + \beta_{hj}\big(E[\tilde{V}^h] - E[V^h]\big) \\ &= b_{ih} \cdot \big[1 + \beta_h\big(E[\tilde{V}^h] - E[V^h]\big)\big] \end{aligned} \tag{20}$$

[15] A deterministic coercion bias would violate this property for interest groups where the support of b_{hj} includes both positive and negative values.

[16] For example, if g_h is the c.d.f. of the uniform distributions, the lower and upper limits would be a function of actual and counterfactual utilities rather than constants.

This implies two things: First, since $\beta_h \geq 0$ and $\tilde{V}^h \geq V^h$, it follows that x_{hj} is at least as dispersed as b_{hj}, implying that $g'_h(-Z_h) \leq f'_h(-Z_h)$. Note also that if voters in interest group h have no coercion bias such that $\beta_h = 0$, (19) reduces to (13). Second, expression (20) implies that $g_h(\cdot) = g_h(-Z_h; E[\tilde{V}^h] - E[V^h])$, and that therefore

$$-\frac{\partial g_h}{\partial E[V_1^h]} = \frac{\partial g_h}{\partial E[\tilde{V}^h]} = \frac{\partial g_h}{\partial\big(E[\tilde{V}^h] - E[V^h]\big)} \tag{21}$$

Using this equality, (19) is equivalent to the FONCs from a Lagrangian that maximizes an additive social welfare function subject to a technology and a coercion constraint of the form

$$\max_t \quad \mathscr{L} = \sum_{h=1}^{H} n_h \left(g'_h + \frac{\partial g_h}{\partial\big(E[\tilde{V}^h] - E[V_1^h]\big)} \right)\bigg|_{t=t_1^*} \cdot V_1^h - \lambda F(\cdot) \tag{22}$$

where the term in parenthesis is constant and can be interpreted as the implicit combined welfare weight for interest group h. As in the case without coercion, the first term corresponds to α_h in the social planner model and measures the density of voters at the policy optimum, the only difference being that g'_h is the p.d.f. of the total bias x_{hj} rather than b_{hj}. The political candidate tailors his policy towards voter types that are more responsive to his policy platform, i.e. swing voters. The second weight measures the marginal loss in vote share due to marginally increasing coercion.[17]

Comparing (22) to the FONCs from the social planner problem make it clear that the effect of including economic coercion is the two approaches is similar if consumers compute correct counterfactuals, because in that case, the derivatives w.r.t. to counterfactual utility drop out in (5). But if consumers make mistakes, the social planner considers the change in counterfactual utility, whereas the political candidate does not.

This is not because the candidate does not know that there will be a change in counterfactual utility (we have assumed throughout that political candidates are fully informed), but because incorporating this anticipated change will decrease his vote share. If voters make mistakes, there would be a different platform that would be overall more satisfying to voters, but this would not become clear until after the election. In this situation, ignoring the change in counterfactuals is a dominant strategy for the political candidate.

To summarize: Whereas the social planner may not give the consumer what he wants

[17] Recall that the probability that a voter in interest group h votes for candidate 1 is $1 - g(\cdot)$, such that $g(\cdot)$ measures that probability that the voter votes for candidate 2.

because he has superior knowledge about consumers' counterfactuals, the political candidate gives voters exactly what they want, even if he knows that this may ultimately not be in their best interest. We refer to this outcome as *counterfactual populism*.

3.3 Incumbent candidates

The previous subsection implicitly assumes that the two candidates want to win the election, but are not concerned with what happens afterwards.

Suppose candidate 1 wins the election on a platform t_1^*, which, as we have seen above, neglects any change in voters' counterfactuals. Once elected, she has to institute an actual policy, and this is where coercion once again comes into play. If coercion of her constituents increases with t_1^* and she has plans to get re-elected after her first term expires, she may want to implement a different policy instead. We will assume that the policy she institutes after being first elected coincides with her policy platform in the second election.[18] Suppressing the expectation signs, voter j in interest group h votes for the incumbent if

$$\begin{aligned} \pi_{1hj} &= 1 \qquad \text{if} \quad V_1^h(t_1) - V_2^h(t_2) - \psi_h\big(V^h(t_1^*) - V_1^h(t_1)\big) > x_{hj} \qquad t_h \geq 0 \qquad (23) \\ \pi_{1hj} &= 0 \qquad \text{otherwise} \end{aligned}$$

where ψ_h captures a change in utility if actual policy t_1 differs from t_1^*. Voters whose utility under the actual policy t_1 is less than under t_1^* are less likely to reelect the incumbent, and vice versa. If $\psi_h = 0$, voters do not care about the politicians' past promises (or do not remember them) but only consider actual welfare. Note that utility under policy t_1^* is a hypothetical utility similar to counterfactual utility, since this policy may have never been in place. If consumers make errors when computing their counterfactual, they may also compute $V^h(t_1^*)$ incorrectly.

Denoting the LHS of (23) as $\bar{Z}_h$ and setting $n_h = 1$ to simplify the notation, the incumbent's vote share maximization problem becomes

$$\max_{t_1} \quad E[VS_1] = \sum_{h=1}^{H} 1 - g_h(-\bar{Z}_h) \qquad s.t. \quad F(\cdot) \leq 0$$

[18] This assumption is based on credibility. If the politician offers a policy other than what is actually observed, then voters will want to know why she did not institute this policy already.

with FONCs of the form

$$
\begin{aligned}
\lambda \frac{\partial F(\cdot)}{\partial t_{1i}} &= \sum_{h=1}^{H} g'_h \frac{\partial \bar{Z}_h}{\partial t_{1i}} - \frac{\partial g_h}{\partial V_1^h}\frac{\partial V_1^h}{\partial t_{1i}} - \frac{\partial g_h}{\partial \tilde{V}^h}\frac{\partial \tilde{V}^h}{\partial t_{1i}} \\
&= \sum_{h=1}^{H} g'_h \left[\frac{\partial V_1}{\partial t_{1i}} - \psi_h \left(\frac{\partial V^h(t_1^*)}{\partial t_{1i}} - \frac{\partial V_1}{\partial t_{1i}}\right)\right] - \frac{\partial g_h}{\partial(\tilde{V}^h - V^h)}\frac{\partial(\tilde{V}^h - V^h)}{\partial t_{1i}} \qquad (24)
\end{aligned}
$$

If voters have no memory of the incumbent's platform when he was running for office for the first time, then $\psi_h = 0$ and the above condition is equivalent the FONCs from the social planner problem in (5). However, if breaking campaign promises is potentially costly such that $\psi_h > 0$, the terms associated with ψ_h have to be considered as well.

We can now state our second result:

Proposition 2. *a.) In two-party competition with probabilistic voting and no incumbent, economic coercion can lead candidates to engage in "counterfactual populism": Even though they recognize that a policy may not be in their constituents' best interest due to its effect on coercion, they nevertheless adopt it as their policy platform because voters' counterfactuals will not change until the policy is instituted. The platform will only reflect the best interest of voters if the latter compute their counterfactual correctly.*

b.) Once elected, economic coercion will lead the incumbent to institute a policy that differs from his election platform, unless voters compute correct counterfactuals or re-election is ruled out. The incumbent weights particular consumer groups negatively along policy dimension t_{1i} if

$$
g'_h < \frac{\frac{\partial g}{\partial \tilde{V}^h}\frac{\partial(\tilde{V}^h - V^h)}{\partial t_{1i}}}{\frac{\partial V^h}{\partial t_{1i}}(1+\psi_h) - \psi_h \frac{\partial V(t_1^*)}{t_{1i}}} \qquad (25)
$$

c.) If voters have no memory of the candidate's platform or do not care about any divergence between actual and promised policy such that $\psi_h = 0\,\forall\, h$, then condition (25) collapses to (9), with $\alpha_h = g'_h$ and $\kappa_h = \partial g_h/\partial V^h$.

Proof. Part a.) follows from the fact that the candidate sets $\partial E[\tilde{V}^h]/\partial t_1 = 0$ in (19) because the counterfactual will not adjust until after the election when policy is set. The second and third points follow from solving the RHS of (24) for g'_h. □

The intuition behind parts b.) and c.) is as follows. Suppose that voters do not punish incumbents for changing their policy relative to their past campaign platform such that $\psi_h = 0$ (this simplifies the argument), and assume further that $\partial V^h/\partial t_{1i} > 0$. Marginally increasing t_{1i} raises the utility and thus increases candidate 1's expected vote share from interest group h. The candidate may want to decrease t_{1i} nevertheless if this increases the election probability enough through the coercion channel. There are two cases where this could occur, provided that the relevant magnitudes are sufficiently large: 1.) Increasing t_{1i} increases coercion, and increased coercion reduces the probability that voters in interest group h vote for candidate 1 (the term $\partial g/\partial(\tilde{V}^h - V^h)$); or 2.) increasing t_{1i} decreases coercion, but a decrease in coercion reduces expected votes.

Increasing coercion leads to a spreading of the bias function. Figure 4 illustrates. The horizontal axis measures $-Z_h$ evaluated at the policy optimum, and the solid line is the c.d.f. of x_{hj} evaluated at $-Z_h$ for a situation with low economic coercion of interest group h. Increasing coercion leads to a spread of the c.d.f., represented by the dotted line. Depending on whether the policy solution lies to the left or to the right of $-\bar{Z}_h$, increasing coercion increases or decreases $g(\cdot)$ (recall that $g(\cdot)$ measures the probability that someone in interest group h votes for candidate 2). In the above case 1, increasing t_{1i} may reduce the expected vote share from this interest group if the solution lies to the left of $-\bar{Z}_h$, or alternatively, if $Z_h|_{t_1^*} > \bar{Z}_h$. This applies to interest groups that on average prefer party 1. The candidate will count their preferences negatively along policy dimension t_{1i} if the effect of attracting further votes from this interest group by increasing their members' utility is more than offset by the effect of driving them towards voting for the other candidate by increasing coercion.

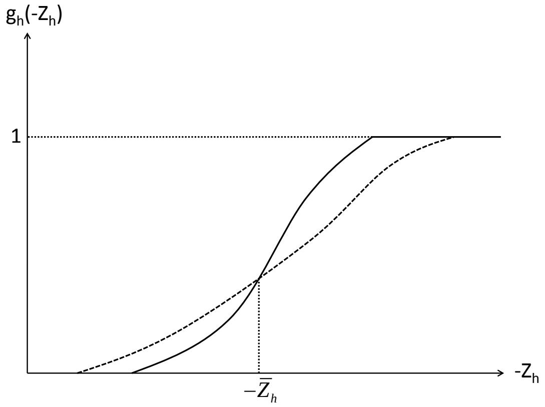

Figure 4: Probability density functions with coercion

Case 2 describes the opposite situation: If increasing t_{1i} reduces coercion but the solution lies to the right of $-\bar{Z}_h$, the positive effect on the vote share by increasing t_{1i} is more than offset by the negative effect through the coercion channel: Decreasing coercion by increasing t_{1i} may make this group more likely to vote for the other candidate.

4 Numerical application

In our numerical application, we choose specific functional forms to represent preferences and technology. We simplify the theoretical model in section 2 by restricting the number of policy instruments to a tax on labor, and a tax on an intermediate good E that is associated with an aggregate environmental externality. The model given here is the non-calibrated version, which is easier for exposition purposes. For the actual solution we transform everything into calibrated share form relative to a benchmark. The exact numeric specification is given as a GAMS code in the Appendix.

4.1 Model

We represent utility by a nested constant elasticity of substitution (CES) function. In the top nest, a private consumption composite C_h and leisure ℓ_h are combined according to a Cobb-Douglas relationship. This private aggregate then produces utility along with public provision G and environmental quality Q:

$$U^h = U^h(C_h, \ell_h, G, Q) = \left[\left(\theta_h^C + \theta_h^\ell \right) \left(C_h^{\delta_h} \ell_h^{1-\delta_h} \right)^{\rho_u} + \theta_h^G G^{\rho_u} + \theta_h^Q Q^{\rho_u} \right]^{1/\rho_u} \tag{26}$$

The parameter δ_h reflects the share in total income spent on private consumption goods within the Cobb-Douglas nest, the θ_h^k, $k = C, \ell, G, Q$, are the share parameters of the CES function with $\sum_k \theta_h^k = 1 \, \forall \, h$, and the exponent $\rho_u \equiv 1 - 1/\sigma_u$ reflects the curvature of the indifference curves, where σ_u is the elasticity of substitution between the private aggregate, public provision and environmental quality.

Consumers maximize their utility subject to the budget constraint

$$M_h = p_L L_h + I_h \leq C_h; \qquad L_h = \bar{L}_h - \ell_h \tag{27}$$

where L_h refers to consumer h's labor supply, p_L is the (uniform) net-of-tax wage and $\bar{L}_h$ represents the time endowment in efficiency units.[19] In addition to income from labor,

[19] This allows us to use one wage rate that applies to all consumers. Consumer heterogeneity is captured by variation in $\bar{L}_h$.

consumers may also receive non-taxable income I_h.[20]

Consumers' shadow value of time is given by ω_h. The comparative slackness condition

$$\omega_h \geq p_L; \quad L_h \geq 0; \quad (\omega_h - p_L) \cdot L_h = 0 \tag{28}$$

allows for the possibility of consumers not entering the labor market, if their valuation of time exceeds p_L. The consumer demand for private consumption and leisure resulting from maximizing (26) subject to (27) is

$$\begin{aligned} C_h &= \frac{\theta_h^C M_h}{p_C} \\ \ell_h &= \frac{\theta_h^L M_h}{p_L} \quad \text{if } \omega_h = p_L; \qquad \ell_h = \bar{L}_h \text{ otherwise} \end{aligned}$$

Production takes place in two stages. In the first stage, labor and I_h are used linearly to produce two intermediate goods:

$$\sum_h (L_h + I_h) = X + E \tag{29}$$

Good X is a clean composite, whereas E is associated with an externality that negatively affects environmental quality according to

$$Q(E) = E^{-\phi} \qquad \phi > 0$$

Due to the linear technology, producer prices for X and E are fixed, and we can choose quantities such that they are unity.[21]

In a second stage, X and E are used to produce final output C and G according to a CES production function:

$$\begin{aligned} f(X, E) = Y &\equiv G + \sum_h C_h \\ &= \Phi\big[tX^{\rho_y} + (1-t)E^{\rho_y}\big]^{1/\rho_y}; \qquad \rho_y = 1 - 1/\sigma_y \end{aligned} \tag{30}$$

Here, $0 \leq t \leq 1$ is a share parameter, σ_y is the elasticity of substitution between X and

[20] The source of this income could be anything, e.g. capital, land or black-market labor; the only restriction is that this income cannot be taxed.

[21] This is the primary reason for defining E as an intermediate input, because it allows for fixed producer prices for E while at the same time not imposing fixed prices for final output.

E in the production of final output Y,[22] and Φ is a constant that will take on the value of expenditure in the benchmark to express utility in money-metric terms.

The government chooses ad-valorem tax rates for labor and the dirty intermediate good, t_L and t_E, respectively, but taxes neither the consumption composite nor the clean intermediate good X. Tax revenue is used to fund the public good:

$$p_Y G = t_L p_L \sum_h L_h + t_E E \tag{31}$$

With CES technology, the price of the final output (which applies to both C and G) is

$$p_Y = \frac{1}{\Phi}\left[t^{\sigma_y} + (1-t)^{\sigma_y}(1+t_E)^{1-\sigma_y}\right]^{1/1-\sigma_y}$$

Demand for the intermediate goods X and E can be derived by solving the cost minimization problem

$$\min_{X,E} \quad X + P_E \cdot E \qquad s.t. \quad f(X,E) \leq Y$$

with resulting demand functions of the form

$$X = \frac{Y}{\Phi} \cdot \left(t p_Y\right)^{\sigma_y}$$
$$E = \frac{Y}{\Phi} \cdot \left(\frac{(1-t)p_Y}{1+t_E}\right)^{\sigma_y}$$

Finally, market clearance implies that

$$Y = \frac{\sum_h \omega_h L_h + I_h}{p_Y} + G$$

Consumers compute their counterfactual level of utility by choosing the taxes on labor and the dirty good. We assume that consumers acknowledge the government's budget constraint (31), but that they only consider partial equilibrium consequences when setting their preferred tax rates.

Let $t = (t_L, t_E)$ denote the vector of tax rates and $z = z(t)$ a vector of producer prices and activity levels. Indirect utility can then be written as $V^h(z)$. The government's

[22] In our numerical model, we choose the same value for the utility function as well as $f(X,E)$, such that $\sigma_u = \sigma_y = \sigma$. Naturally, this is a special case.

coercion-constrained welfare maximization problem can be written abstractly as[23]

$$\max_t \sum_h \alpha_h V^h(z) \qquad s.t. \quad F(z,t) = 0 \tag{32}$$

$$t_h\big(\tilde{z}_h, \tilde{t}_h; z, t\big) = 0 \tag{33}$$

$$V^h(z) \geq (1 - \bar{v}) \cdot \tilde{V}^h(\tilde{z}) \tag{34}$$

where (32) represents all general equilibrium constraints (technology, environmental damages and market clearance, represented by eqs. (29-31), (33) is a constraint requiring joint consistency of realized and counterfactual equilibria for all consumer groups, and (34) is a coercion constraint stipulating that actual welfare must not fall below a fraction of counterfactual utility, with $0 < \bar{v} < 1$. The α_h's are individual welfare weights.

We solve the model in calibrated share form and compute welfare, prices and quantities relative to a benchmark defined by $t_L = 0.5$, $t_E = 1$ (i.e. a 50% tax on labor, and a 100% tax on the dirty good). We separate consumers into three distinct groups:

1.) The "Green Young" have a high labor endowment $\bar{L}_h$ but little other income I_h and represent the workforce. They have a relatively low preference for the public good, but a high preference for environmental quality.

2.) The "Green Old" represent a consumer group with high non-labor income I_h and a high preference for environmental quality and the public good, but which does not participate in the labor market (technically, their shadow value of time exceeds the wage rate due to a low labor endowment)

3.) The "Brown Old" are similar to the Green Old except that they do not value environmental quality as highly.

4.2 Results

Figure 5 shows preferred policy gradients at the benchmark in two-dimensional policy space defined by the levels of the tax on labor and on the dirty good. Since all tax revenue is used to finance the single public good, the level of the latter is implicit and increases towards the upper right of the figure.

In keeping with their endowments and preferences, the Green Young prefer a lower

[23] In the numerical application, we work with calibrated share forms, which are less parsimonious than the model representation here. The GAMS code containing the full model specification is in the appendix.

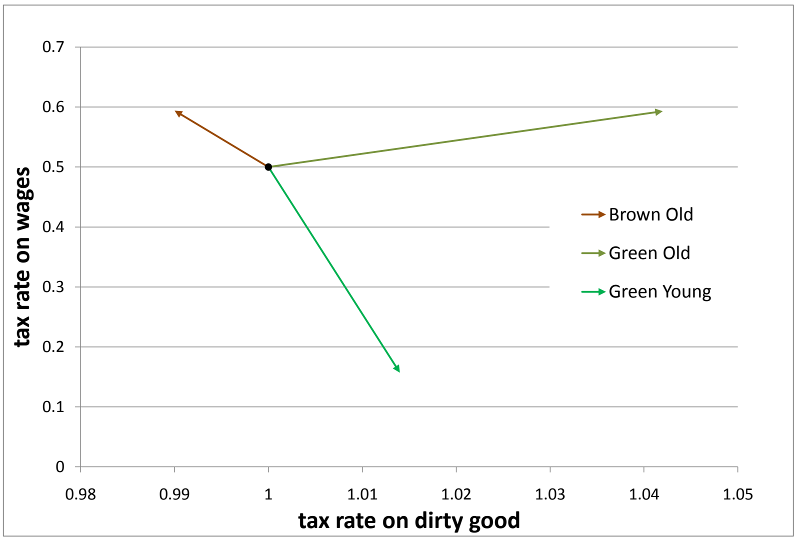

Figure 5: Preferred policy gradients in $t_E - t_L$ space

labor tax but a higher tax on the dirty good, relative to the benchmark. The Green Old would like to increase both taxes (which increases the level of the public good), whereas the Brown Old would prefer a higher tax on labor but but a lower tax on the dirty good (because a high tax on E lowers the price for both factors, including I_h). The figure implies that any change in policy away from the benchmark can be obtained by a change in the relative utility weights α_h that the social planner places on the three groups.

Figure 6 shows the optimal (general equilibrium, or GE) policy choices along with the corresponding counterfactual (partial equilibrium, or PE) solutions for the three groups, along with policy "approach paths" (solid for GE, dashed for PE). We derive the policy approach paths in the following manner: Starting from the benchmark represented by the black dot, we draw a circle around it and compute the optimal point on that circle for each group according to the GE and the PE representation of the model. This is equivalent to maximizing welfare for the group in question while including a constraint that the solution has to be within a given distance of the benchmark. We then increase the radius of the circle and repeat this calculation, until we reach the equilibrium points.

The figure displays two features of our model, both of which are related to the type of error we assume. First, the difference between the PE and GE paths increases as we move away from the benchmark. Consumers have rational counterfactuals as long as they have knowledge of the equilibrium (which is the case at the benchmark), but as we move away their utility depends on demand and supply levels that they miscalculate by ignoring general equilibrium effects. Second, the counterfactual choices involve higher tax rates

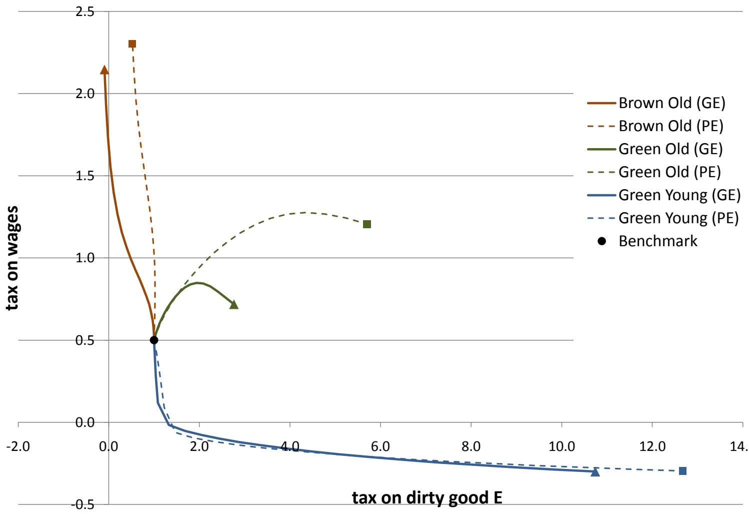

Figure 6: Optimal and counterfactual policy choices ($v_G = v_Q = 3$)

than the optimal levels for all consumer groups. Neglecting general-equilibrium effects of taxation underestimates the total distortion introduced by a single tax, a well-known result of the optimal taxation literature.

Last, the figure shows that the approach paths are nonlinear. This is true for all groups, but it is particularly visible for the Young Green. Close to the benchmark, both the actual and perceived optimal policy change consist in reducing the labor tax. Once t_L reaches zero, an increase in t_E becomes the desired policy change. The intuition behind is that being the only workers in the economy, the Young Green have a strong incentive to decrease the wage tax, with the effect of increasing output and their purchasing power and at the cost of a reduction in the level of public provision (for which they have a low valuation). But in order to actually receive a wage subsidy, the dirty good must be taxed highly, since this is the only other source of government income.

Figure 6 is based on medium tastes for public provision (v_G) and environmental quality (v_Q).[24] In comparison, Figures 7 and 8 show the policy solutions for the cases of high v_G / low v_Q, and the reverse situation, respectively. The increased valuation of public provision or environmental quality leads to higher optimal tax rates, but also to an increase in the counterfactual error.

Figure 9 shows how the counterfactual error changes with the equilibrium from which they are computed. We divide the transition between the benchmark and the true opti-

[24] These are input parameters reflecting the utility derived from the public good and environmental quality at the benchmark. Dividing these utilities by the benchmark welfare (which is in money-metric terms) yields the share parameters θ_h^G and θ_h^Q in the utility function (26).

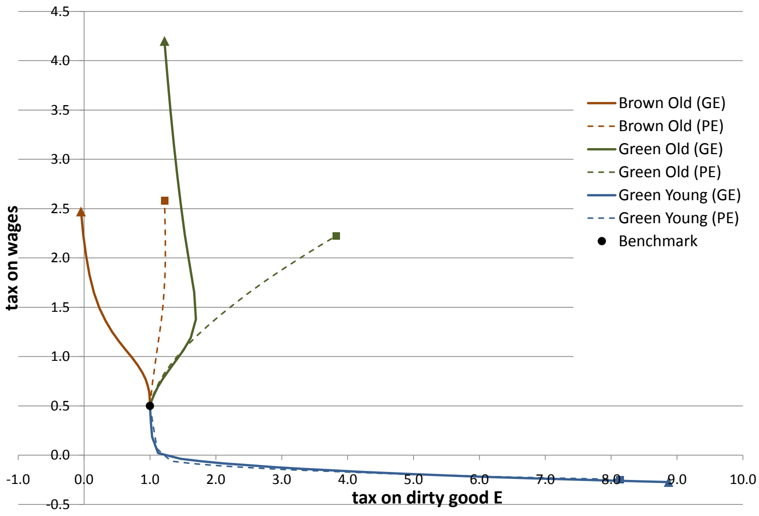

Figure 7: Optimal and counterfactual policy choices ($v_G = 4, v_Q = 2$)

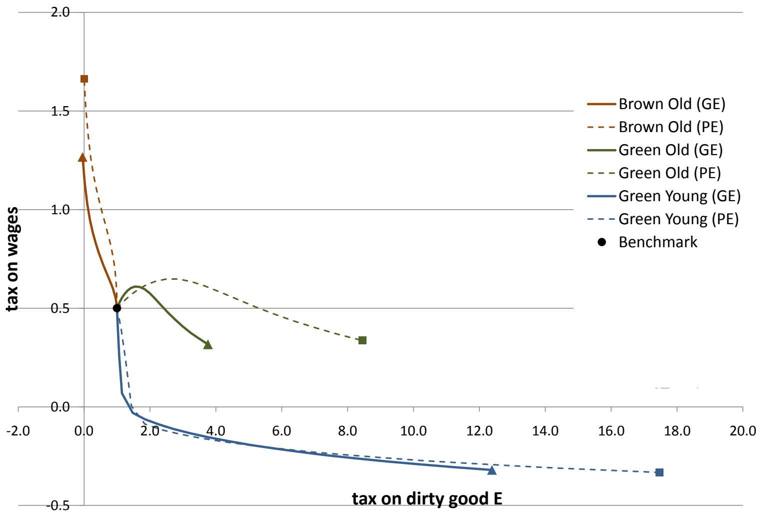

Figure 8: Optimal and counterfactual policy choices ($v_G = 2, v_Q = 4$)

mum into 20 steps, where step 0 represents the benchmark, and step 20 is the group GE optimum (the triangles in the figures). The figure shows the counterfactual optima computed at steps 0, 5, 10, 15 and 20. Counterfactual policies evaluated at the groups' GE optima do not coincide with these optima; in other words, consumers prefer a different policy even at the policy that maximizes their true utility, because they (wrongly) believe that a different point in policy space would be better for them. Even moving to that particular counterfactual point would not remove economic coercion completely, since the counterfactual would again shift.[25]

[25] For every consumer group, there exists a policy point that coincides with consumers' counterfactuals. We have not finished the algorithm to find these fixed points, but hope to do so in a future version of this manuscript.

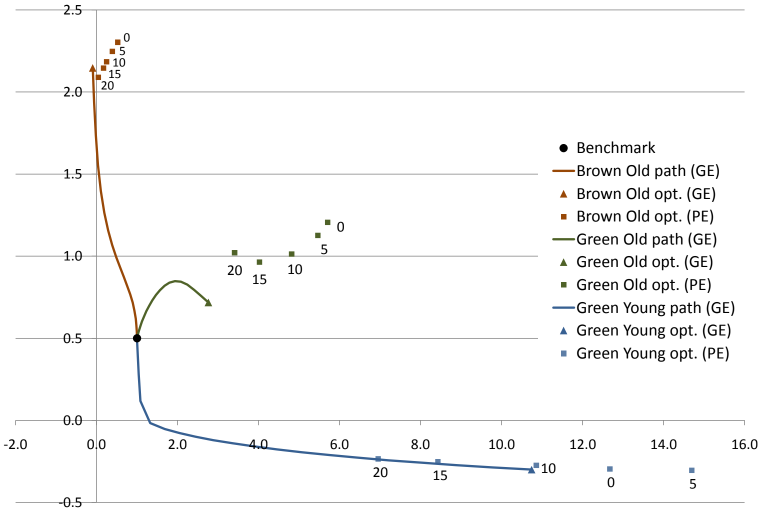

Figure 9: Dependency of counterfactuals from actual equilibria

5 Conclusions

Introducing a limit on how much people may be coerced by society's rules can be justified for several reasons, such as the protection of minorities and the wish to preempt socially damaging action from individuals who feel disenfranchised. While governments will always employ negotiations to settle labor disputes and law enforcement to prevent and punish illegal activities, mitigating disenfranchisement by limiting economic coercion may be a valid additional strategy.

The effect of limiting economic coercion depends crucially on whether consumers/voters make mistakes when computing their counterfactual utilities. If they are completely rational and fully informed, they will solve the same general equilibrium model as the social planner, in which case their counterfactual utility becomes a constant. The model can then be rewritten by replacing the coercion constraint with additional welfare weights for economically coerced consumers. These additional weights would have to be calculated based on the level of economic coercion, but formally the model corresponds to "traditional" social planning, and the outcome will always be allocationally efficient.

However, if consumers are not able to solve the full general equilibrium model of a modern economy, their counterfactuals will be endogenous to current policy. In this case, we cannot simply place additional welfare weight on economically coerced groups, because the additional terms differ by policy dimension. This is easiest to accommodate by addressing coercion in the model as an explicit constraint, even though the resulting equilibrium may be replicated by some different set of welfare weights, which are a complicated function of weighted coercion effects across the various policy dimensions. If the counterfactuals are sufficiently wrong and the coercion constraint sufficiently tight, the policy outcome may even be inconsistent with nonnegative welfare weights and thus be allocationally inefficient. We derive a necessary condition for this to occur.

Limiting economic coercion if consumers are not fully rational and/or informed can be interpreted as a particular form of paternalism, which we refer to as counterfactual paternalism. By this we mean that there exist policy choices that would increase overall welfare and which the coercion-constrained consumer group deems desirable based on a counterfactual evaluated at the coercion-constrained solution. However, changing policy in this direction would lead this consumer group to adjust their counterfactual in a way that leads to an increase in economic coercion, and thus the violation of the coercion constraint.

Similarly, a political candidate will consider the endogeneity of voters' counterfactuals, but he will use this knowledge in order to win an election as opposed to protecting minorities or reduce social strife. Because counterfactuals will not adjust until a policy is implemented, economic coercion provides an incentive for a politician to institute a policy that differs from the pre-election platform. We refer to this situation as counterfactual populism.

Applying our model to actual data could be a fruitful avenue for future work. In a normative context, our framework could guide policy makers in the design of policies that limit the extent of economic coercion to particular groups at the lowest cost to overall welfare. Furthermore, if incorrect counterfactuals have a significant effect on policy outcomes and thus on welfare this would be a reason to design institutions with the aim of reducing counterfactual errors. Lastly, our model could be used in a positive context to explain observed policy choices.

References

Baumol, W. (2003) “Welfare Economics and the Theory of the State,” *Encyclopedia of Public Choice*, Vol. 2, pp. 610–613. edited by Charles K. Rowley and Fridrich Schneider.

Bernheim, B. Douglas and Antonio Rangel (2007) “Behavioral Public Economics: Welfare and Policy Analysis with Nonstandard Decision-Makers,” in Peter Diamond and Hannu Vartiainen eds. *Behavioral Economics and its Applications*: Princeton University Press, p. 7.

Buchanan, James M. and Gordon Tullock (1962) *The Calculus of Consent. Logical Foundations of Constitutional Democracy*: University of Michigan Press.

Gamson, William A. (1961) “A Theory of Coalition Formation,” *American Sociological Review*, Vol. 26, No. 3, pp. 373–382.

Lindahl, Eric (1919) “Just Taxation, a Positive Solution,” in R. Musgrave and A. Peacock eds. *Classics in the Theory of Public Finance (1958)*: MacMillan, p. 168.

Mueller, Dennis C. (2003) *Public Choice III*: Cambridge University Press. p. 253-257.

Skarpedas, Stergios (1992) “Cooperation, Conflict, and Power in the Absence of Property Rights,” *American Economic Review*, Vol. 82, No. 4, pp. 720–739.

Wicksell, Knut (1896) “A New Principle of Just Taxation,” in R. Musgrave and A. Peacock eds. *Classics in the Theory of Public Finance (1958)*: MacMillan, p. 72.

Winer, S., G. Tridimas, and W. Hettich (2008) “Social Welfare and Coercion in Public Finance.” CESifo Working Paper Series No. 2482.