Designing a leaky bucket: Meltzer and Richard (1981) with endogenous inefficiency in redistributive institutions

Ahmet Faruk Aysan*and Haldun Evrenk†

May 9, 2015

Abstract

Unlike what Meltzer and Richard (1981) assume, and due to inefficient in-kind transfers, the administrative costs of enforcement and redistribution, the tax payers cost of compliance and non-compliance, what tax payers paid for income redistribution is never completely redistributed. Extending their model, we study the relationship between inequality and redistribution when there are endogenously determined inefficiencies (leakage) in redistributive institutions. The level of leakage is decided on in a constitutional period by a designing voter that ranks higher than median. By increasing the cost of redistribution, the leakage reduces future median's demand for equilibrium redistribution; it also increases the incentives to work and the future income. In societies with high (low) levels of initial inequality, the designing voter sets inefficient (efficient) redistributive institutions. Since the past inequality reduces the demand for redistribution (through leakage) and current inequality increases the demand for redistribution, the net effect of inequality on redistribution is not clear. This provides an explanation for the weak correlation between income inequality and equilibrium redistribution obtained in many cross-country regressions.

Key Words: Inequality, efficient redistribution, institutions, political economy, endogenous labor supply

"The money must be carried from the rich to the poor in a leaky bucket. Some of it will simply disappear in transit, so the poor will not receive all the money that is taken from the rich." Okun (1975, p. 91)

"Constitution was calculated to increase the influence, power and wealth of those who have any already." John Quincy Adams

*Central Bank of the Republic of Turkey, Ankara, Turkey.

†Economics Department, TOBB ETU, Sogutozu Cad. No 43. Ankara, Turkey. Email: haldun_evrenk@yahoo.com

1 Introduction

In their seminal model of political economy of income redistribution, Meltzer and Richard (1981), hereafter MR, assume that the total payment by taxpayers is exactly equal to the total income (re)distributed. The actual redistribution, however, is never that efficient. Jacoby (1997), for example, notes that a free lunch supplement program in Jamaica is valued at around \$158 by the households, while it costs \$400 to the state. In addition to inefficient in-kind transfers, the administrative cost of enforcement and redistribution, the taxpayers' cost of compliance when the tax code is complex as well as their cost of non-compliance (bribes and fees) when evasion or avoidance is possible, all contribute to the discrepancy between the cost of income redistribution to the taxpayers and its benefit to those who receive it. Below, we refer to this discrepancy as leakage (and, sometimes as the inefficiency in redistributive institutions).

In this paper, we first extend the model in MR by adding some exogenous leakage. Then, we further extend it by endogenizing the leakage. Under an exogenous level of leakage, we find that (by increasing the cost of redistribution) leakage reduces equilibrium redistribution. Yet, the existence of leakage does not necessarily change the main prediction of MR, the mean-to-median-income-ratio (MMIR) hypothesis, in a cross country regression. If all countries have the same level of leakage, the model with leakage, too, predicts that the countries with higher levels of inequality (as measured by the pre-tax MMIR) will redistribute a higher fraction of their aggregate income. Since the anecdotal evidence suggests that the level of leakage varies significantly among countries, we endogenize the level of leakage by assuming that the (efficiency of) redistributive institutions were set up in a constitutional period (or, when transitioning to democracy) by a *designing* voter. With this second extension a possible theoretical explanation for the weak empirical support for the MMIR hypothesis emerges.[1]

We assume that the designing voter ranks higher than the median in the initial income (or, productivity, as the wage is the only source of income in MR) distribution.[2] The designing voter can set neither the fully inefficient institutions nor the perfectly efficient institutions. The latter is not available, while the former is not sustainable, –always a majority will support a reform targeting such a large degree of inefficiency. But, in between these two extremes, there exist a continuum of alternatives for the designing voter to choose from. Under the simplifying assumption that the individual elasticities of labor supply are constant, we find that

[1] See Borck (2007) for a review of some of these empirical studies. The lack of evidence also could be due to problems with the data employed. Instead of the post-redistribution incomes commonly used, Milanovic (2000) uses pre-tax incomes and finds weak support for the MMIR, –but employing similar data and a non-parametric regression, Pecoraro (2014) fails to find any support. Another data problem is that due to non-uniform participation in elections the pre-tax income level of the actual median voter (and, thus, the MMIR) is difficult to identify.

[2] See, Engerman and Sokoloff (1997), Acemoglu, Johnson and Robinson (2000) and the following literature for a similar approach to institutions.

the designing voter's optimization problem always has a corner solution. If his productivity is higher (lower) than a threshold, the designing voter will set up the most (least) inefficient redistributive institution available (Proposition 1). The intuition for this result is that, leakage reduces both how much tax payers pay for income redistribution and how much they receive as the result of redistribution. The cost (lower demogrant) is the same for everyone but the benefit (lower taxes) increases in the productivity (or, wage income) of the designing voter. For the voter with the threshold productivity, the cost and benefit are equal to each other.

To see how this result could explain the lack of evidence for MMIR hypothesis, consider two societies h and l, with (resp.) high and low initial income inequality. If in both societies the designing voter ranks the same (above the median, but not too much) in the productivity distribution, then in l (h) his productivity would be below (above) the threshold, so l ends up with efficient redistributive institutions and redistribute a higher fraction of its income than h which ends up with highly inefficient redistributive institutions and little redistribution. Therefore, in a cross section of countries with sufficient variation in initial inequality, we expect at least two[3] groups of countries with differing levels of leakage, and within each group the MMIR hypothesis to hold. Yet, when one pools data from two or more groups without any control for leakage, the extended MR model studied here has no clear prediction on the correlation between inequality and redistribution.

Our result that countries with initial economic or political inequality may end up with highly inefficient redistributive institutions is relevant not just for the weak empirical evidence for MMIR hypothesis. Several issues in the literature are studied under the assumption that the state redistributes in the most efficient way possible. For example, while discussing whether the distributional aspects of the law should be kept in mind when the laws are designed, Kaplan and Shavell (1994) argue that when the income is redistributed through the income tax there is only a single distortion, i.e., the state redistributes in as efficiently as possible, while the using the law to redistribute would lead to a "double distortion." Likewise, in his discussion of the effectiveness of constitutional constraints, Evrenk (2009) assumes that if such constraints are not imposed, then the political actors with the power to impose these constraints will not use any other instrument to limit the size or the growth of the government. If the alternative to constitutional constraints are redistributive institutions with built-in inefficiency, however, the constraints may be welfare improving.

Even the specific literature on inequality, redistribution and institutions is extensive. Here we briefly discuss previous studies particularly related to general income redistribution with endogenously determined inefficiency in redistributive institutions.[4] Closest to our work, Aysan

[3] More, if for different countries the least and the most efficient institutions available were different.

[4] Dixit and Londregan (1995; 1996), Acemoglu and Robinson (2001), and Becker and Mulligan (2003) are also related, but they focus on inefficiencies in *targeted redistribution* (price supports and subsidies to inefficient sectors). These are outside of the scope of the MR model and, thus, our extension.

(2005) provides several explanations for the lack of empirical support for the MMIR hypothesis, including endogenously determined inefficiency of redistributive institutions in a setup with income maximizing voters and a log-normal income distribution. Using theoretical analysis instead of numerical simulations, our paper provides a more formal framework for that explanation.

Chong and Gradstein (2007), too, study the relationship between income redistribution, inequality, and institutions. They study a dynamic model in which the labor supply is exogenous and income is redistributed through rent seeking (from poor to rich). Inefficient institutions increase the return from rent seeking and benefit richer individuals at the expense of the poor. The quality of institutions in their model is voted on in every period. Their decisive voter is richer than the median due to political bias. Despite several differences in the modelling assumptions, the authors, too, find that the institutional quality converges to the lowest (highest) possible levels of efficiency with high (low) initial inequality. They also present empirical evidence on the persistence of both inequality and low institutional quality, even though these institutions are not necessarily redistributive ones.

Acemoglu, Ticchi, and Vindigni (2011) show how public good production can be manipulated through patronage politics, and their model can easily be interpreted as a model of income distribution (utility is linear in both public and private good consumption). They consider a society with three groups of people, poor, rich, and the bureaucrats; the collected tax revenue depends on how hard bureaucrats work, while the labor supply of the rest is exogenous. With sufficient inequality, there is an equilibrium in which the state collects little tax revenue (and, thus, redistributes little), supported by a coalition of the bureaucrats and the rich as both parties benefit from the low effort by bureaucrats. If one assumes that the rich and the bureaucrats are numerous enough to form a majority, then the idea of patronage politics in Acemoglu, Ticchi, and Vindigni (2011) can explain inefficiency in government as the equilibrium outcome of one period political competition without any need for institutions. However, both anecdotal and empirical evidence suggest that especially when transititioning to democracy limits to redistribution were set.

Karl (1990) mentions the "elite pacts" in several Latin American countries, setting hurdles against redistribution when transitioning to democracy. There, too, is a converse relationship between inequality and redistribution: the higher is the inequality, the higher is the protection against redistribution that the elite would demand to allow a transition to democracy. Illustrating this point with a global data set, Albertus and Menaldo (2014) show that if the elites are politically weak during the transition to democracy, then there is a positive correlation between democracy and redistribution. They also discuss many ways in which elites bias democratic institutions.

2 The model

There is a continuum of individuals of measure one, indexed by their productivity $x \in \mathbb{R}_+$. Their distribution is characterized by the c.d.f. $F(x)$. Each individual has preferences over two normal goods: a composite consumption good, c, and leisure, l. These preferences are represented by a continuous, twice differentiable, and strictly concave utility function $U(c, l)$. Each individual is being paid according to his productivity, working n units of time, the individual with productivity level x earns a pretax income of $y(x) = xn$. Then he pays an income tax at the flat rate t, and receives the demogrant R from the collected tax revenue. Consumption is equal to the disposable income, $(1-t)xn + R$. The time endowment of each individual is normalized to one; thus, leisure is given by $l = 1 - n$.

For given t and R, an individual with productivity level x chooses n to maximize $u((1-t)xn + R, 1-n)$, leading to the first order condition

$$u_c(\tau xn + R, 1-n)\tau x - u_l(\tau xn + R, 1-n) \leq 0 \quad (= 0 \text{ if } n > 0), \tag{1}$$

where

$$\tau = 1 - t.$$

Equation (1) implicitly defines the labor supply of the individual with productivity x as a function of the demogrant and after-tax wage rate, $n(R, (1-t)x)$. It also follows from (1) that an individual will not work (subsist on payment R) if and only if his productivity level is less than the threshold

$$x_0 = \frac{u_l(R, 1)}{u_c(R, 1)(1-t)}. \tag{2}$$

Thus, the aggregate (and, average) pre-tax income is equal to

$$\overline{y} = \int_{x_0}^{\infty} xn(R, (1-t)x)dF(x), \tag{3}$$

As in MR, we impose the following assumption.

Assumption 1 (*Constant Aggregate Elasticities*): The partial elasticities of aggregate income with respect to R and τ are constant with

$$\frac{\partial \overline{y}}{\partial R}\frac{R}{\overline{y}} = \eta_R \text{ and } \frac{\partial \overline{y}}{\partial \tau}\frac{\tau}{\overline{y}} = \eta_\tau \text{ for all } R \text{ and } \tau.$$

We use η_R and η_τ instead of $\eta(\overline{y}, R)$ and $\eta(\overline{y}, \tau)$ used in MR, to simplify the notation. More important, we use R (not r) to denote the demogrant, because we consider the possibility of inefficiency in redistribution (or, leakage). Formally, we have

$$R = (1-\lambda)t\overline{y},$$

where $\lambda \in [0,1)$ denotes the level of leakage. Since there is no leakage in MR (they have $r = t\overline{y}$), the model in MR is a special case of the model we study with $\lambda = 0$.

One can interpret λ as the fraction of resources wasted due to inefficient in-kind redistribution, a measure of administrative inefficiency, the deadweight loss of tax evasion and avoidance, or the cost of tax compliance. To emphasize the latter two interpretations, one could rewrite the model with $R = t\overline{y}$, while assuming that a taxpayer gives up effectively a $\frac{t}{1-\lambda}$ fraction of his pre-tax income when the official tax rate is t. None of our results would change under that alternative formulation.

To solve for the equilibrium tax rate and redistribution level for a given λ, we follow the same steps as in MR, –see Meltzer and Richard (1981, p. 920-22).[5] We can simply follow these steps because introducing the leakage to the model does not violate the single-crossing property of the voter preferences. That is, the slope of a voter's indifference curve in (R,t) space (v_t/v_R) is still equal to $xn(x)$, where $v(.)$ is the indirect utility function that follows from (1). Since that slope is an increasing function of productivity, the existence of a unique voting equilibrium is guaranteed. The equilibrium tax rate is the one most preferred by the median voter.

Let x_d denote the voter with the median productivity (and, thus, with the median income), and let $m = \overline{y}/y_d$ denote the ratio of the mean income to the median income.[6] Then the most preferred tax rate of the median voter is

$$t = \frac{(1-\lambda)m - 1 + \eta_R}{(1-\lambda)(1+\eta_\tau)m - 1 + \eta_R}. \tag{4}$$

From (4), we obtain the following comparative statics.

Lemma 1 *(i) For a given level of leakage, equilibrium tax rate and fraction of total income redistributed increases in the mean-to-median income ratio (MMIR),* $\frac{\partial t}{\partial m} = \frac{(1-\lambda)(1-\eta_R)\eta_\tau}{((1-\lambda)(1+\eta_\tau)m-1+\eta_R)^2} > 0$.

(ii) For a given MMIR, equilibrium tax rate and fraction of total income redistributed decreases in leakage, $\frac{\partial t}{\partial \lambda} = -\frac{(1-\eta_R)m\eta_\tau}{((1-\lambda)(1+\eta_\tau)m-1+\eta_R)^2} < 0$ *and* $\frac{\partial (R/\overline{y})}{\partial \lambda} = -t + (1-\lambda)\frac{\partial t(\lambda,m)}{\partial \lambda} < 0$.

Part (i) of Lemma 1 implies that in a cross country regression, the extended model, too, gives rise to the MMIR hypothesis when all countries have similar levels of leakage. If, as the anecdotal evidence suggests, λ is not the same in every country, then this is not true anymore:

[5] To save space we do not report detailed calculations here. They are available upon request.

[6] As in MR, we focus on the case in which $x_d > x_0$ (the median voter works) and $m > 1$.

inefficiency in redistribution reduces the median voter's demand for redistribution (Part (ii) of Lemma 1). Next, we endogenize the level of leakage.

To endogenize the level of leakage, we assume that the (efficiency of) redistributive institutions are determined by a *designing* voter in a constitutional period. The designing voter may be different from the median voter. To simplify the calculations we assume that he is able to set long lasting rules and regulations (institutions), and he is able to calculate the effects of these on equilibrium tax rate and redistribution in a deterministic world.[7]

Formally, he solves

$$\max_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} u^D(R(\lambda, m(\lambda)) + x_D n(\lambda, x_D)(1 - t(\lambda, m(\lambda)), 1 - n(\lambda, x_D)), \tag{5}$$

where $\underline{\lambda} > 0$ and $\overline{\lambda} < 1$ denote, respectively, upper and lower bounds on the leakage. Such bounds are reasonable, because a minimum level of leakage may be unavoidable due to administrative costs and redistributive institutions that are too inefficient may not survive the future attempts to reform.

Using the Envelope Theorem and (1), one can show that the marginal utility from the leakage for the designing voter is given by

$$\frac{dU_D(.)}{d\lambda} = (-\frac{dt}{d\lambda} y_D + \frac{dR}{d\lambda}) \frac{\partial U_D(.)}{\partial c}, \tag{6}$$

where $y_D = x_D n(\lambda, x_D)$ is the future income of the designing voter.

We know that the marginal utility from consumption is positive, therefore to sign $\frac{dU_D(.)}{d\lambda}$, we need to sign $-\frac{dt}{d\lambda} y_D + \frac{dR}{d\lambda}$. For this, note that

$$\frac{dt}{d\lambda} = \frac{\partial t}{\partial \lambda} + \frac{\partial t}{\partial m} \frac{dm}{d\lambda} \tag{7}$$

$$\frac{dR}{d\lambda} = -t\overline{y} + (1 - \lambda) t \frac{d\overline{y}}{d\lambda} + (1 - \lambda) \overline{y} \frac{dt}{d\lambda} \tag{8}$$

$$\frac{dm}{d\lambda} = \frac{1}{y_d} \frac{d\overline{y}}{d\lambda} - m \frac{1}{y_d} \frac{dy_d}{d\lambda} \tag{9}$$

$$\frac{d\overline{y}}{d\lambda} = \frac{\partial \overline{y}}{\partial R} \frac{dR}{d\lambda} - \frac{\partial \overline{y}}{\partial \tau} \frac{dt}{d\lambda} \tag{10}$$

$$\frac{dy_d}{d\lambda} = \frac{\partial y_d}{\partial R} \frac{dR}{d\lambda} - \frac{\partial y_d}{\partial \tau} \frac{dt}{d\lambda} \tag{11}$$

[7] All of these assumptions can be reasonably weakened without any qualitative effect on our main result.

Solving these five equations, and manipulating the solutions, we find

$$\frac{dt}{d\lambda} = \frac{-m\eta_\tau(1-t(1-\lambda)my_{dR})}{((1-\lambda)(1+\eta_\tau)m - 1 + \eta_R)^2 - (1-\lambda)\eta_\tau(1-\eta_R)((\overline{y}_R - my_{dR}) - \frac{1}{y_d}(\overline{y}_\tau - my_{d\tau}))} \tag{12}$$

$$\frac{dR}{d\lambda} = \frac{-t\overline{y}}{1-\eta_R} + \frac{dt}{d\lambda}y_d \tag{13}$$

$$\frac{dm}{d\lambda} = \frac{-tm}{1-\eta_R}\frac{\eta_R - \varepsilon_{Rd}}{(1-\lambda)t} + \frac{dt}{d\lambda}(\frac{\eta_R - \varepsilon_{Rd}}{(1-\lambda)t} - \frac{(\eta_\tau - \varepsilon_{\tau d})m}{1-t}) \tag{14}$$

where $y_{dR} = x_d\frac{\partial n(R,x_d\tau)}{\partial R}$, $y_{d\tau} = x_d\frac{\partial n(R,x_d\tau)}{\partial \tau}$, $\varepsilon_{Rd} = \frac{\partial n(R,x_d\tau)}{\partial R}\frac{R}{n(R,x_d\tau)}$ and $\varepsilon_{\tau d} = \frac{\partial n(R,x_d\tau)}{\partial \tau}\frac{1-t}{n(R,x_d\tau)}$.

Without further restriction, we cannot determine the sign of $-\frac{dt}{d\lambda}y_D + \frac{dR}{d\lambda}$ in (6), and, thus, cannot identify the solution to the optimization problem in (5). If, in addition to the *Constant Aggregate Elasticities* assumption in MR, we assume that the individual elasticities are constant, then we have a well defined solution. Formally, our assumption is as follows.[8]

Assumption 2 (*Constant Individual Elasticities*): The elasticities of individual labor supply are constant for working population. That is, for all $n(R, x\tau) > 0$, we have

$$\frac{\partial n(R,x\tau)}{\partial R}\frac{R}{n(R,x\tau)} = \varepsilon_R \text{ and } \frac{\partial n(R,x\tau)}{\partial \tau}\frac{1-t}{n(R,x\tau)} = \varepsilon_\tau.$$

Assumption 2 allows us to solve the optimization problem for the designing voter because imposing a restriction on individual elasticities while there is already a restriction on aggregate elasticities has strong implications on who works and how much. For individual elasticities to be constant while the aggregate elasticities are constant, a change in the tax rate or in the demogrant must affect how much an individual works, but not *whether* he works. Then, the set of people who work does not depend on R and τ (but, *how much* they work does).[9] Thus, if the aggregate income changes at a constant rate, so should the aggregate labor supply: the latter is simply the former multiplied by the productivities of the workers that participate in the labor force. But, then the set of workers and, thus, the mean-to-median income ratio are independent of the degree of leakage.[10] Formally,

[8] *Constant Individual Elasticities* assumption differs from the *Constant Aggregate Elasticities* in two dimensions: individual vs. aggregate and income vs. labor supply. The second difference is artificial. *Constant Individual Elasticities* can be written in terms of pre-tax income instead of individual labor supply. To see why, simply note that the elasticities of pre-tax income of a working individual i can be written as $\frac{\partial y_i(R,x_i\tau)}{\partial R}\frac{R}{y_i(R,x_i\tau)} = \frac{\partial n(R,x_i\tau)}{\partial R}\frac{R}{n(R,x_i\tau)}$ and $\frac{\partial y_i(R,x\tau)}{\partial \tau}\frac{1-t}{y_i(R,x\tau)} = \frac{\partial n(R,x\tau)}{\partial \tau}\frac{1-t}{n(R,x\tau)}$, where $y_i(R, x_i\tau) = x_i n(R, x_i\tau)$.

[9] So, this is a strong assumption. In its defense we can say what MR states in defense of their *Constant Aggregate Elasticities* assumption: we expect that our results will still hold when the individual elasticities are not constant, but have limited variation.

[10] Even though m does not change, the income inequality as measured by any other measure that satisfies Dalton Principle, e.g. Gini Coefficient, will. As λ increases such measures will find higher income inequality. This is because, under *Constant Individual Elasticities* the set of working people do not change in λ, while we do have $\frac{d\overline{y}}{d\lambda} > 0$, –see the proof of Proposition 1 for more on that.

Lemma 2 *When both the aggregate and individual elasticities are constant, each of the individual elasticities is equal to the corresponding aggregate one ($\eta_R = \varepsilon_R$ and $\eta_\tau = \varepsilon_\tau$). Moreover, an increase in leakage changes the pre-tax incomes, but not their ratios. Particularly, we have $\frac{d(y_D/y_d)}{d\lambda} = 0$ and $\frac{dm}{d\lambda} = 0$.*

Lemma 2 is proved in the Appendix. Using it, we solve (5) and establish our main result.

Proposition 1 *There exists a threshold productivity level x^* such that for any $x_D < (>) x^*$, the designing voter chooses the lowest (highest) possible leakage he can. This productivity level lies strictly between the mean (among the working population) and median productivity ($\int_{x_0}^{\infty} x dF(x) > x^* > x_d$).*

The proof of Proposition 1 is provided in the Appendix. To summarize it, *Constant Individual Elasticities* is a sufficient condition for both the tax rate and the demogrant to be decreasing in leakage when the income is endogenous.[11] Thus, an increase in leakage brings a benefit (lower taxes) at the cost of a lower demogrant. Then, by (6), any level of leakage will be desirable for a designing voter with sufficiently low productivity (income), and, it will be undesirable for a designing voter with sufficiently high productivity (income). For productivity levels in between, the marginal utility from leakage is first decreasing then increasing. This is because, due to lower taxes and demogrant the designing voter's income increases in leakage, $\frac{dy_D}{d\lambda} > 0$. Thus, for designing voters whose productivity is neither too high nor too low, the utility function is U-shaped in leakage, and the optimal level of leakage depends on whether $u^D(\overline{\lambda})$ is larger than $u^D(\underline{\lambda})$. Furthermore, the difference between $u^D(\overline{\lambda})$ and $u^D(\underline{\lambda})$ decreases in x_D, with $u^D(\overline{\lambda}) = u^D(\underline{\lambda})$, at $x^D = x^*$. Therefore when the designing voter's productivity level is exactly x^*, he is indifferent between $\overline{\lambda}$ and $\underline{\lambda}$, but this is a non-generic case.

Note that other than saying it is above the median productivity and below the average productivity among the working population, we cannot identify the productivity threshold x^* unless we have the functional form for the utility (and, thus, solve for x^D that makes $u^D(\overline{\lambda}) = u^D(\underline{\lambda})$). Yet, what we want to emphasize is the implications of the existence of such a threshold, and not the actual level of the threshold. This is because modifying some of the modelling assumptions with other ones, e.g., the designing voter facing some uncertainty about his future productivity, would change the threshold but not that implication.

To see the implication of Proposition 2 that can explain the lack of empirical evidence for MMIR hypothesis in cross–country regressions, consider two countries, i and j, where the productivity distribution in j has been obtained by moving some of the mass just above the median in i to both below and well above the median. Thus, productivity distribution

[11] For the case with exogenous income, Lemma 1 already states that result. Also note that even for the endogenous income case, the *Constant Individual Elasticities* assumption is much stronger than necessary.

in country j has a lower median and a higher average. Assume that the set of available institutions for both countries are the same, i.e., $[\underline{\lambda}_i, \overline{\lambda}_i] = [\underline{\lambda}_j, \overline{\lambda}_j]$. Also assume that the designing voter comes from the same percentile of productivity distribution in both countries, and that percentile is above the fifty percent (the median) but not too much. By choosing that percentile for the designing voter's productivity appropriately we can have x_D to be above the threshold in Proposition 2 in j while it is below the corresponding threshold in i.[12] Then, country i will end up with better institutions ($\underline{\lambda}_i$), and j will end up with the worse institutions, $\overline{\lambda}_j$. If the difference between these two types of redistributive institutions (the distance $\overline{\lambda}_j - \underline{\lambda}_i$) is sufficiently large, then the low inequality country i will redistribute a higher fraction of its income in equilibrium.

The above example illustrates how endogenous institutions may help explain the lack of empirical support for the MMIR hypothesis in cross-country regressions. It also illustrates that endogenously determined institutions do not necessarily lead to a failure of the MMIR hypothesis; we need additional assumptions. We discuss these briefly below.

First, the relative ranking (of the productivity, and, thus, income) of the designing voter must be above the median but not too much above it. Because, only then the productivity of the designing voter would fall above (below) of the threshold productivity when the level of inequality is high (low). If the designing voter's productivity always falls on the same side of x^* (if he is always the richest person, or if he is always the median), then independent of the degree of inequality, the leakage will be the same throughout the world. Then, by Lemma 1, the model studied here, too, would always give rise to the MMIR hypothesis.

Second, and related to the first point, in our argument it is implicitly assumed that the designing voters in different societies will have the same rank (are from the same percentile) in the income distribution. That particular assumption is not necessary to make our point. All we need is that the relative ranking of the designing voter should not be strongly and *negatively* correlated with income inequality. Because, if the designing voters in countries with high (low) inequality were relatively poor (rich) then country i (j) in the above example would choose the most (least) efficient redistributive institutions, further strengthening the predictions of MR. This assumption can be justified as follows as well. As long as we let the designing voter to be someone with above the median income, we are effectively allowing political power to depend on economic power when it comes to setting up the institutions.[13] Then, it is reasonable to expect that the relative ranking of the designing voter to increase as the income inequality

[12] The corresponding threshold will be different for i, but in the direction desirable for our argument: in i the median is higher and the average is lower.

[13] This could happen, for instance if the voters with low income (productivity) do not have the right to vote or do not have the necessary information to choose the best institutions for their own economic interest, when the institutions were designed.

increases.

Third, we assume that the available institutions for the designing voter are identical in different countries. What matters is that the least efficient institution available in (the high inequality country) j is not more efficient than the most efficient institution available in (the low inequality country) i. Otherwise, the designing voter in j will choose the least efficient institution, but that will be still more efficient than the most efficient institution in i. Apart from such drastic variations in the institutions available, the available lower or the upper bounds for the efficiency are likely to be different in different countries.[14] As a result, there will be several (not just two) sets of countries with different leakage levels. Our analysis predicts that within each set the MMIR will hold, but comparing two countries from different sets, it may or it may not.

3 Conclusion

In this paper we explore the implications of endogenous inefficiency in redistributive institutions within the framework of the standard models of income redistribution in Meltzer and Richard. The predictions of the extended model studied for cross-country regressions are more nuanced. An appropriate test of our model requires identifying the level of inefficiency in redistributive institutions in different countries. Due to the lack of reliable and comprehensive data on the efficiency of redistributive institutions for a large set of countries, this is left for future research.

4 Appendix

Proof of Lemma 2. Since $n(R, x_0\tau) = 0$, by Leibnitz' rule, we have $\overline{y}_R = \int_{x_0} x \frac{\partial n(R,x\tau)}{\partial R} dF(x)$ and $\overline{y}_\tau = \int_{x_0} x \frac{\partial n(R,x\tau)}{\partial \tau} dF(x)$. Using the definitions of $\varepsilon_R = \frac{\partial n(R,x\tau)}{\partial R} \frac{R}{n(R,x\tau)}$ and $\varepsilon_\tau = \frac{\partial n(R,x\tau)}{\partial \tau} \frac{1-t}{n(R,x\tau)}$, we can rewrite these as

$$\overline{y}_R = \int_{x_0} x \varepsilon_R \frac{n(R, x\tau)}{R} dF(x) = \varepsilon_R \frac{\overline{y}}{R} \tag{15}$$

$$\overline{y}_\tau = \int_{x_0} x \varepsilon_\tau \frac{n(R, x\tau)}{1-t} dF(x) = \varepsilon_\tau \frac{\overline{y}}{1-t} \tag{16}$$

Using equations (18) and (19) in the definitions of aggregate elasticities, $\eta_R = \overline{y}_R \frac{R}{\overline{y}}$ and $\eta_\tau = \overline{y}_\tau \frac{1-t}{\overline{y}}$, we have

$$\eta_R = \varepsilon_R \text{ and } \eta_\tau = \varepsilon_\tau. \tag{17}$$

Next, we show that as λ changes, the ratio $\frac{y_D}{y_d}$ remains unchanged, i.e., $\frac{d(y_D/y_d)}{d\lambda} = 0$. Note

[14] Because, for example, the variations in the designing voter's political power, –see Albertus and Menaldo (2014).

that
$$\frac{d(y_D/y_d)}{d\lambda} = (\frac{\frac{dn(R,x_D\tau)}{d\lambda}n(R,x_d\tau) - \frac{dn(R,x_d\tau)}{d\lambda}n(R,x_D\tau)}{(n(R,x_d\tau))^2})\frac{x_D}{x_d},$$
where $\frac{dn(R,x_D\tau)}{d\lambda}n(R,x_d\tau) - \frac{dn(R,x_d\tau)}{d\lambda}n(R,x_D\tau) = (\frac{\partial n(R,x_D\tau)}{\partial R}\frac{dR}{d\lambda} - \frac{\partial n(R,x_D\tau)}{\partial \tau}\frac{dt}{d\lambda})n(R,x_d\tau) - (\frac{\partial n(R,x_d\tau)}{\partial R}\frac{dR}{d\lambda} - \frac{\partial n(R,x_d\tau)}{\partial \tau}\frac{dt}{d\lambda})n(R,x_D\tau)$. Rearranging this as $\frac{dR}{d\lambda}(\frac{\partial n(R,x_D\tau)}{\partial R}n(R,x_d\tau) - \frac{\partial n(R,x_d\tau)}{\partial R}n(R,x_D\tau)) - \frac{dt}{d\lambda}(\frac{\partial n(R,x_D\tau)}{\partial \tau}n(R,x_d\tau) - \frac{\partial n(R,x_d\tau)}{\partial \tau}n(R,x_D\tau))$, and using the definitions of ε_R and ε_τ, we have $\frac{dR}{d\lambda}(\varepsilon_R\frac{n(R,x_D\tau)}{R}n(R,x_d\tau) - n(R,x_D\tau)\varepsilon_R\frac{n(R,x_d\tau)}{R}) - \frac{dt}{d\lambda}(\varepsilon_\tau\frac{n(R,x_D\tau)}{1-t}n(R,x_d\tau) - \varepsilon_\tau\frac{n(R,x_d\tau)}{1-t}n(R,x_D\tau)) = 0$. Thus,
$$\frac{d(y_D/y_d)}{d\lambda} = 0.$$

Finally, we show that $\frac{dm}{d\lambda} = 0$. Note that, using (9), (10), and (11), we have
$$\frac{dm}{d\lambda} = \frac{1}{y_d}((\frac{\partial \overline{y}}{\partial R} - m\frac{\partial y_d}{\partial R})\frac{dR}{d\lambda} - (\frac{\partial \overline{y}}{\partial \tau} - m\frac{\partial y_d}{\partial \tau})\frac{dt}{d\lambda}) \tag{18}$$
Using (18), (19), and rewriting $m\frac{\partial y_d}{\partial R}$ as $mx_d\varepsilon_R\frac{n(R,x_d\tau)}{R}$, and $m\frac{\partial y_d}{\partial \tau}$ as $mx_d\varepsilon_R\frac{n(R,x_d\tau)}{R}$, we have $\frac{dm}{d\lambda} = 0$. ■

Proof of Proposition 1. Using Lemma 2, we can simplify (12) as $\frac{dt}{d\lambda} = \frac{-m\eta_\tau(1-\eta_R)}{((1-\lambda)m(1+\eta_\tau)-1+\eta_R)^2} < 0$, and, thus, $\frac{dU_D(.)}{d\lambda}$ as
$$\frac{dU_D(.)}{d\lambda} = E\overline{y}\frac{\partial U_D(.)}{\partial c}, \tag{19}$$
where
$$E = (\frac{-1}{1-\eta_R}\frac{(1-\lambda)m - 1 + \eta_R}{(1-\lambda)m(1+\eta_\tau) - 1 + \eta_R} + (\frac{y_D}{y_d} - 1)\frac{\eta_\tau(1-\eta_R)}{((1-\lambda)m(1+\eta_\tau) - 1 + \eta_R)^2}). \tag{20}$$
Note that the sign of $\frac{dU_D(.)}{d\lambda}$ depends only on the sign of E. Also note that, by Lemma 2, the ratio $\frac{y_D}{y_d}$ does not change in λ, it only depends on the ratio of the productivities of designing and the median voters.

When $\frac{y_D}{y_d} \leqslant 1$, at every λ, m, η_τ, and η_R we have $E < 0$, and, thus, $\frac{dU_D(.)}{d\lambda} < 0$. Therefore, when the designing voter's productivity is less than the median, he sets $\lambda = \underline{\lambda}$.

When $\frac{y_D}{y_d} > 1$, there are three cases. First, for $\frac{y_D}{y_d} > \frac{((1-\underline{\lambda})m(1+\eta_\tau)-1+\eta_R)((1-\underline{\lambda})m-1+\eta_R)}{\eta_\tau(1-\eta_R)^2} + 1$, we have $\frac{dU_D(.)}{d\lambda} > 0$ for all λ. Therefore, the designing voter sets $\lambda = \overline{\lambda}$. Second, for $\frac{y_D}{y_d}$ sufficiently small, $\frac{y_D}{y_d} < \frac{((1-\overline{\lambda})m(1+\eta_\tau)-1+\eta_R)((1-\overline{\lambda})m-1+\eta_R)}{\eta_\tau(1-\eta_R)^2} + 1$, we have $\frac{dU_D(.)}{d\lambda} < 0$ for all λ. Therefore, the designing voter sets $\lambda = \underline{\lambda}$.

Third, for $\frac{y_D}{y_d}$ between the two bounds mentioned above, i.e.,
$$\frac{((1-\overline{\lambda})m(1+\eta_\tau) - 1 + \eta_R)((1-\overline{\lambda})m - 1 + \eta_R)}{\eta_\tau(1-\eta_R)^2} < \frac{y_D}{y_d} - 1 < \frac{((1-\underline{\lambda})m(1+\eta_\tau) - 1 + \eta_R)((1-\underline{\lambda})m - 1 + \eta_R}{\eta_\tau(1-\eta_R)^2} \tag{21}$$

we show that $U_D(\lambda)$ is first decreasing and then always increasing. For this case, by continuity we have $\frac{dU_D(.)}{d\lambda} \geq 0$ at some λ_0 (otherwise, the optimal λ would simply be $\underline{\lambda}$). Since the sign of $\frac{dU_D(.)}{d\lambda}$ depends only on the sign of E, and $E > 0$ is a *sufficient* condition for $\frac{dE}{d\lambda} > 0$, for the same individual $\frac{dU_D(.)}{d\lambda} > 0$ for all $\lambda \in (\lambda_0, \overline{\lambda})$. Likewise, for λ small enough, we have $\frac{dU_D(.)}{d\lambda} < 0$. Thus, $U_D(\lambda)$ is U shaped: its first decreasing, then increasing. Therefore, in the third case the designing voter will set $\lambda = \overline{\lambda}$ when $u^D(\overline{\lambda}) > u^D(\underline{\lambda})$, and $\lambda = \underline{\lambda}$ when $u^D(\overline{\lambda}) < u^D(\underline{\lambda})$, where $u^D(\lambda) = u^D(R(\lambda) + x_D n(\lambda, x_D)(1 - t(\lambda), 1 - n(\lambda, x_D)))$.

Let x^* denote the productivity level for the designing voter under which $u^D(\overline{\lambda}) = u^D(\underline{\lambda})$. Since $\frac{dU_D(.)}{d\lambda}$ is a continuous and increasing function of λ and y_D (and, thus, x_D), we know that x^* exists and it is unique. Then, we have the desired result.

Above, we find that the threshold productivity is above the median. Now, we show that it is below the mean. As (6) shows, this depends on how the average after tax income, $(1-t)\overline{y}+(1-\lambda)t\overline{y} = (1-\lambda t)\overline{y}$, changes in λ. Differentiating w. r. to λ, we have $\frac{d[(1-\lambda t)\overline{y}]}{d\lambda} = (1-\lambda t)\frac{d\overline{y}}{d\lambda} - t\overline{y} - \lambda\overline{y}\frac{dt}{d\lambda} = (1-\lambda t)(\frac{\partial\overline{y}}{\partial R}\frac{dR}{d\lambda} - \frac{\partial\overline{y}}{\partial\tau}\frac{dt}{d\lambda}) - t\overline{y} - \lambda\overline{y}\frac{dt}{d\lambda} = (1-\lambda t)(\frac{\eta_R}{(1-\lambda)t}\frac{dR}{d\lambda} - \frac{\eta_\tau\overline{y}}{1-t}\frac{dt}{d\lambda}) - t\overline{y} - \lambda\overline{y}\frac{dt}{d\lambda}$. Using (13), we have, $\frac{d[(1-\lambda t)\overline{y}]}{d\lambda} = t\overline{y}(\frac{(1-\lambda t)}{(1-\lambda)t}\frac{-\eta_R}{1-\eta_R} - 1) + \frac{dt}{d\lambda}\overline{y}(\frac{(1-\lambda t)\eta_R}{(1-\lambda)mt} - \frac{(1-\lambda t)\eta_\tau}{1-t} - \lambda)$. Note that $\frac{dt}{d\lambda} < 0$ and $\frac{(1-\lambda t)\eta_R}{(1-\lambda)mt} - \frac{(1-\lambda t)\eta_\tau}{1-t} - \lambda < 0$, thus, if $\frac{(1-\lambda t)}{(1-\lambda)t}\frac{-\eta_R}{1-\eta_R} - 1 \geq 0$, then we have $\frac{d[(1-\lambda t)\overline{y}]}{d\lambda} > 0$. So, we only need to sign $\frac{d[(1-\lambda t)\overline{y}]}{d\lambda}$ for the case $\frac{(1-\lambda t)}{(1-\lambda)t}\frac{-\eta_R}{1-\eta_R} - 1 < 0$. That is, for $\frac{-\eta_R}{(1-\lambda)} < \frac{t}{1-t}$. The solution to the median voter's optimal tax rate problem implies $\frac{t}{1-t} = \frac{1}{\eta_\tau}(1 - \frac{(1-\eta_R)}{(1-\lambda)m})$. Thus, $\frac{-\eta_R}{(1-\lambda)} < \frac{t}{1-t} = \frac{1}{\eta_\tau}(1 - \frac{(1-\eta_R)}{(1-\lambda)m})$. Therefore, the set of parameters we are looking for can be characterized as

$$-\eta_R m\eta_\tau < (1-\lambda)m - (1-\eta_R).$$

For this case, $\frac{d[(1-\lambda t)\overline{y}]}{d\lambda} = \overline{y}[\frac{(1-\lambda t)(-\eta_R)-(1-\lambda)t(1-\eta_R)}{(1-\lambda)(1-\eta_R)} + \frac{dt}{d\lambda}(\frac{(1-\lambda t)\eta_R}{(1-\lambda)mt} - \frac{(1-\lambda t)\eta_\tau}{1-t} - \lambda)]$, where $\frac{dt}{d\lambda} = \frac{-m\eta_\tau(1-\eta_R)}{((1-\lambda)m(1+\eta_\tau)-1+\eta_R)^2}$ under the Constant Individual Elasticities Assumption. Using the fact that $t = \frac{(1-\lambda)m-1+\eta_R}{(1-\lambda)(1+\eta_\tau)m-1+\eta_R}$, and simplifying, further, we have
$\frac{d[(1-\lambda t)\overline{y}]}{d\lambda} = \frac{\overline{y}}{(1-\eta_R)}\frac{((1-\lambda)m-(1-\eta_R))(\eta_\tau(1-\eta_R)^2(m-1)-1)+m\eta_\tau(\eta_\tau(1-\eta_R)^2((1-\lambda)m-1)-\eta_R)}{((1-\lambda)m(1+\eta_\tau)-1+\eta_R)}$.

Then, all we need to show is that

$$((1-\lambda)m - (1-\eta_R))(\eta_\tau(1-\eta_R)^2(m-1)-1) > -m\eta_\tau\eta_R(\frac{\eta_\tau(1-\eta_R)^2((1-\lambda)m-1)}{\eta_R} - 1).$$

As noted above, we are studying the case in which $((1-\lambda)m - (1-\eta_R)) > -m\eta_\tau\eta_R$. For $\eta_\tau(1-\eta_R)^2(m-1) - 1 > \frac{\eta_\tau(1-\eta_R)^2((1-\lambda)m-1)}{\eta_R} - 1$, we need $(m-1) > \frac{(1-\lambda)m-1}{\eta_R}$. Since $\eta_R < 0$, this holds. ■

References

[1] Acemoglu, D., Johnson, S., and Robinson, J. A. (2001) The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review, 91, 1369-1401

[2] Acemoglu, D. and Robinson, J. A. (2001) Inefficient Redistribution. The American Political Science Review, 95: 649-661.

[3] Acemoglu, D., Ticchi, D., and Vindigni, A. (2011) Emergence and Persistence of Inefficient States. Journal of the European Economic Association, 9(2), 177-208.

[4] Albertus, M., and Menaldo, V. (2014) Gaming Democracy: Elite Dominance During Transition and the Prospects for Redistribution. British Journal of Political Science, 44: 575-603.

[5] Aysan, A. F. (2005) Inequality, Institutions and Redistribution. PhD Dissertation. University of Maryland, College Park.

[6] Becker, G. and Mulligan, C. B. (2003) Deadweight Costs and the size of the government. Journal of law and Economics, 46: 293-340.

[7] Borck, R. (2007) Voting, inequality, and redistribution. Journal of Economic Surveys, 21: 90-109.

[8] Chong, A. and Gradstein, M. (2007) Inequality and Institutions. The Review of Economics and Statistics, 89(3): 454-465.

[9] Dixit, A. and Londregan, J. (1995) Redistributive Politics and Economic Efficiency. American Political Science Review 89:856-66.

[10] Dixit, A., and Londregan, J. (1996) The Determinants of Success of Special Interests in Redistributive Politics. Journal of Politics 58: 1132-55.

[11] Engerman, S.L. and K.L. Sokoloff (1997) Factor endowments, institutions, and differential growth paths among new world economies." in S. Haber, ed., How Latin America Fell Behind, Stanford, CA: Stanford University Press, pp. 260–296.

[12] Jacoby, H. G. 1997. Self-selection and the redistributive impact of in-kind transfers: an econometric analysis. Journal of Human Resources. 32, 233-249.

[13] Karl, T. L. (1990) Dilemmas of Democratization in Latin America, Comparative Politics, 23: 1-21.

[14] Kaplow, L. and Shavell, S. (1994) Why the Legal System Is Less Efficient than the Income Tax in Redistributing Income, Journal of Legal Studies, 23: 667-689.

[15] Meltzer, A. H. and Richard, S. F. (1981) A Rational Theory of the Size of Government. Journal of Political Economy, 89: 914-27.

[16] Milanovic, B. (2000) The median-voter hypothesis, income inequality, and income redistribution: an empirical test with the required data, European Journal of Political Economy, 16: 367-410

[17] Okun, A. 1975. Equality and Efficiency, the big tradeoff. The Brooking Institution, Washington, DC.

[18] Pecaro, B. (2014) Inequality in democracies: Testing the classic democratic theory of redistribution, Economics letters, 123: 398-401.

[19] Reinikka, R. and Svensson, J. (2004) Local Capture: Evidence from a Central Government Transfer Program in Uganda, The Quarterly Journal of Economics, 119: 679-705.

[20] Roberts, K. (1977) Voting over income tax schedules. Journal of Public Economics Vol 8, 329-40.

[21] Romer, T. (1975) Individual welfare, majority voting and the properties of a linear income tax, 7: 163-68.