Inequality-Minimization with a Given Public Budget

Johannes König
Carsten Schröder

School of Business & Economics
Discussion Paper
Economics
2016/16
Inequality-Minimization with a Given Public Budget

Johannes König⋆, Carsten Schröder♠

Abstract. We solve the problem of a social planner who seeks to minimize inequality via transfers with a fixed public budget in a distribution of exogenously given incomes. The appropriate solution method depends on the objective function: If it is convex, as in the case of the absolute mean deviation, it can be solved by an interior-point algorithm. If it is quasiconvex, as in case of the Gini coefficient, the bisection method can be used. We implement the procedures using artificial and real-world data, and show that the optimal transfer scheme need not comply with a transfer scheme that perfectly equalizes incomes at the bottom of the distribution.

Keywords: redistribution, public transfers, inequality, optimization methods, interior-point algorithm, bisection method

JEL-codes: C61, D63, H23, H53, I38

⋆: Freie Universität Berlin, School of Business and Economics, j.koenig@fu-berlin.de
♠: DIW Berlin/SOEP & Freie Universität Berlin, School of Business and Economics, cschroeder@diw.de

We thank Giacomo Corneo and Shlomo Yitzhaki for valuable discussions.
1. Introduction

All over the world, policy-makers seek to reduce inequalities in market incomes via transfers (Joumard et al. (2012)). Their distributional effects – their progressiveness – are usually evaluated by comparing inequality indices from the market and post-transfer distributions. Unanswered is the issue of the effectiveness of the transfer scheme: Is there an alternative scheme - for a given public budget - that yields a stronger inequality reduction, and what is the maximum feasible inequality reduction? The present paper shows that the maximum inequality reduction need not imply income equalization at the bottom of the distribution, and identifies optimal transfer schemes by solving constrained minimization problems.

In our context, the problem is to minimize the inequality in a distribution of exogenously given incomes by means of non-negative transfers with a fixed public budget. If the objective function, the index of interest, and the set of constraints is convex, the program can be solved by an interior-point algorithm. If the objective function is quasiconvex and the set of constraints is convex, the program can be solved by the bisection method. Convexity requires that the second derivative of the objective function is non-negative. Quasiconvexity requires that the function’s sublevel sets are convex. For example, the variance is convex and the Gini coefficient is quasiconvex. At the heart of every solution to a minimization problem is the descent along the negative gradient of the objective function. The effect of a marginal monetary transfer to any given household is determined by the first derivative of the objective function with respect to the income of the recipient: the larger the first derivative in absolute terms, the larger the inequality reduction.

What determines the first derivative? Suppose the population is homogeneous, meaning all units are of equal composition, thus having the same material needs. The first derivative of any standard inequality measure increases in the income or rank of the transfer recipient. This means transfers should always be donated to the household units with the smallest pre-transfer income.

As an example, suppose the planner’s objective is the minimization of the Gini index, the public budget is 1 monetary unit, and the income distribution is (14, 20, 30, 40, 80) with a Gini index of 0.33043. The optimal post-transfer distribution is (15, 20, 30, 40, 80) with a Gini index of 0.32432. If the budget were 20 units, the optimal post-transfer distribution would be (27.5, 27.5, 30, 40, 80) with a Gini index of 0.22927. The latter distribution is optimal as the entire budget is channeled to the households at the bottom of the distribution and marginal social utilities for all transfer recipients (first derivatives of the Gini w.r.t. transfers) are equal. We call the underlying transfer rule “bottom fill-up,” as it minimizes income differences at the bottom of the distribution.

Now suppose the population is heterogeneous, meaning household units differ in composition (i.e., number of household members) and needs, with the latter being measured by an equivalence scale. The common practice of measuring inequality in a heterogeneous population involves two steps. The first step is the needs adjustment of incomes by dividing a household unit’s income by its equivalence scale. The second step involves the weighting of household units to construct the equivalent-income distribution. The traditional approach in inequality measurement is to weight households by the number of household members (size weighting). It is consistent with the welfarist’s principle of normative individualism: each person is as important as any other. Then the first derivative of the objective function is not determined by income rank alone. Consider again the aforementioned income distribution (14, 20, 30, 40, 80) but now suppose

1An equivalence scale measures household-size economies and differences in needs across household members (e.g. of adults and children). For example, the square-root equivalence scale adjusts household income using the square root of the number of household members.
household sizes are $\{1, 4, 1, 4, 1\}$. Under size-weighting, the sorted equivalent income distribution under the square-root scale is $\{(10, 4)\ (14, 1), (20, 4), (30, 1), (80, 1)\}$, with the first number giving equivalent income and the second the unit’s weight. The resulting Gini index is 0.36215. For the public budget of 20 units, the bottom fill-up rule gives $\{(18, 4)\ (18, 1), (20, 4), (30, 1), (80, 1)\}$ with a Gini of 0.24026. There is, however, a transfer scheme with an alternative feasible post-transfer distribution $\{(17, 1), (18.5, 4), (20, 4), (30, 1), (80, 1)\}$ that results in a lower Gini coefficient of 0.23940. Bottom fill-up fails to produce the optimal distribution because, for the size-weighted heterogeneous distribution, the inequality-reducing effect of transfers not only depends on (a) the recipient household’s rank in the equivalent income distribution (as in case of a homogeneous population), but also on (b) the household’s weight, and (c) the transfer-induced change in average, equivalent income (which again depends on the recipient household’s weight and equivalence scale). The bottom fill-up rule ignores channels (b) and (c).

The basic intuition why bottom fill-up fails for a size-weighted distribution is the following: Transferring income not to the poorest but to a less poor and larger household unit can imply a stronger increase of average equivalent income. This effect reduces any scale-invariant inequality index, and can overcompensate the potential inequality reduction that could have been achieved by shrinking income gaps at the very bottom of the distribution.

Section 2 shows solutions to the aforementioned optimization problem for convex and quasi-convex inequality indices using constrained optimization techniques. Section 3 provides empirical applications. Section 4 concludes.

2. The Constrained Optimization Problem

Following Ebert and Moyes (2003), suppose households are defined by two attributes: household income $y_i \geq 0$ and household type $h_i \in H=\{1,\ldots,H\}$. Household material well-being is defined by equivalent income, the ratio of household income and the household’s equivalence scale $ES_i = ES(h_i) > 0$. In total, the population consists of $N$ households and $Q = \sum_{i=1}^{N} q_i$ individuals, with $q_i$ denoting the number of individuals in household $i$. Let $w_i$ denote the weight of a household. In case of a homogeneous or size-weighted heterogeneous population $w_i = q_i$. In case of a needs-weighted heterogeneous population $w_i = ES_i$. Average equivalent income is $\bar{y} = (\sum_{i=1}^{N} w_i (y_i + t_i) / ES_i) / W$ with $W = \sum_{i=1}^{N} w_i$.

The aim of the social planner is to minimize inequality for a given income distribution, $Y$, via transfers, $t_i \geq 0$, with a given public budget, $B = \sum_{i=1}^{N} t_i$. The planner assigns the same weight to each individual of every household (size-weighting). Inequality is inferred from the distribution of equivalent income - household income, $y_i$, divided by equivalence scale, $ES_i$. The planner’s objective function is an inequality

\[\text{Under size weighting, economies of scale create a wedge between household size and needs (equivalence scale). A possible way to avoid the wedge is to abandon the principle of normative individualism, and weight households by needs rather than size. Specifically, under needs weighting, the transfer-induced change of average equivalent income does not depend on the recipient household’s composition. Characterizations of needs weighted distributions can be found in the theoretical works of Ebert (1999), Ebert and Moyes (2003), and Shorrocks (2004). The downside is an “ethical dilemma because individuals who have less extensive needs would be given a lower weight” (Wodon and Yitzhaki (2005, p. 3)).}\]

\[\text{Extension to the case of the } t_i \geq 0 \text{ is straightforward, which enables a full derivation of a tax and transfer system. Further, the investigator may introduce distortions. In the presence of such distortions, } \delta_i, \text{ the transfer net of the distortion is } \tilde{t}_i = t_i - \delta_i \text{ and the post-transfer income is } y_i - \delta_i + t_i. \text{ The public budget is } B = \sum_{i=1}^{N} t_i.\]
index, $I : R^N \to R$, which is at least twice continuously differentiable and convex or quasiconvex. Thus the optimization problem is,

$$\min_{t_i} I \left( \left\{ \frac{w_i + ty_i}{ES_i} \right\}_{i=1}^N \right),$$  \hspace{1cm} (1)

subject to inequality constraints,

$$0 \leq t_i \ \forall i,$$ \hspace{1cm} (2)

and

$$\sum_{i=1}^N t_i \leq B.$$ \hspace{1cm} (3)

This implies the following Karush-Kuhn-Tucker (KKT) conditions for optimality (first-order and slack conditions,

$$\frac{\partial L}{\partial t_i} = \frac{dI}{dt_i} \left( \left\{ \frac{w_i + ty_i}{ES_i} \right\}_{i=1}^N \right) - \lambda + \nu_i = 0 \implies \lambda - \nu_i \forall i$$ \hspace{1cm} (4)

and

$$\sum_{i=1}^N t_i^* \leq B, \ 0 \leq t_i^* \ \forall i, \ \nu_i \geq 0, \ \nu_i(-t_i^*) = 0,$$

where $\lambda$ is the Lagrange multiplier associated with the budget constraint and the $\nu_i$ are associated with the non-negativity constraints of the $t_i$. If the $\nu_i$ are zero and, consequently not binding at the optimum, then the marginal social utilities of all transfer recipients should be equal to the shadow-price $\lambda$. Take, for example, the Gini index,

$$G = \frac{2}{1/W} \sum_{i=1}^N w_i \left( \frac{y_i + t_i}{ES_i} \right) - \bar{y} \left( F \left( \frac{y_i + t_i}{ES_i} \right) - \bar{F} \right),$$ \hspace{1cm} (5)

where $F(\cdot)$ is the weighted cumulative distribution function (CDF).\(^6\)

\(^4\)Shalit and Yitzhaki (2005) minimize Gini's mean difference subject to linear constraints in a finance context, while Yitzhaki (1982) minimizes the squares of the differences of pre- and post-reform after-tax income subject to non-linear constraints. Both papers consider a convex objective function, while we also provide solutions for quasiconvex functions.

\(^5\)See Luenberger (1968) and Arrow and Enthoven (1961) for references on quasiconvex programming.

\(^6\)We treat $F(\cdot)$ as twice differentiable, with first derivative denoted by $f(\cdot)$ and second derivative $f'(\cdot)$. The mean of the CDF is $\bar{F} = 1/W \sum_{i=1}^N w_i F \left( \frac{y_i + t_i}{ES_i} \right)$. For the empirical application we use a modification in the formula of the Gini coefficient to make it differentiable. See Appendix B.
Let \( u = \frac{1}{W} \sum_{i=1}^{2} w_i \) and \( v = \frac{1}{W - W/N} \sum_{i=1}^{N} w_i \left( \frac{y_i + t_i}{ES_i} - \bar{y} \right) \left( F \left( \frac{y_i + t_i}{ES_i} \right) - \bar{F} \right) \). The first-order condition is,

\[
\frac{\partial G}{\partial t_i} - \lambda = \frac{\partial u}{\partial t_i} v + \frac{\partial v}{\partial t_i} u - \lambda = 0, \tag{6}
\]

with

\[
\frac{\partial u}{\partial t_i} = -2Ww_i/ES_i \]

and

\[
\frac{\partial v}{\partial t_i} = \frac{1}{W - W/N} \left[ \frac{w_i}{ES_i} \left( (F - \bar{F}) + \left( \frac{y_i + t_i}{ES_i} - \bar{y} \right) f \left( \frac{y_i + t_i}{ES_i} \right) \right) - \sum_{j=1}^{N} \frac{w_jw_i}{ES_iW} \left( (F - \bar{F}) + \left( \frac{y_j + t_j}{ES_j} - \bar{y} \right) f \left( \frac{y_j + t_j}{ES_j} \right) \right) \right].
\]

The derivative \( \frac{\partial u}{\partial t_i} \) reflects the marginal effect of a transfer on average equivalent income. The derivative \( \frac{\partial v}{\partial t_i} \) shows that the transfer’s effect on inequality depends on the recipient household’s weight and equivalence scale \( \frac{w_i}{ES_i} \), together with its position in the equivalent income distribution.

For a homogeneous population

\[
\frac{\partial u}{\partial t_i} = -\frac{2}{N\bar{y}^2},
\]

\[
\frac{\partial v}{\partial t_i} = \frac{1}{N-1} \left[ \left( (F - \bar{F}) + \left( (y_i + t_i) - \bar{y} \right) f \left( y_i + t_i \right) \right) - \sum_{j=1}^{N} \frac{1}{N} \left( (F - \bar{F}) + \left( (y_j + t_j) - \bar{y} \right) f \left( y_j + t_j \right) \right) \right].
\]

and \( \frac{\partial \bar{y}}{\partial t_i} = \frac{\partial \bar{y}}{\partial t_j} \); hence, the redistributive effect of \( t_i \) depends on the household’s position in the equivalent income distribution but not on the household’s weight (channel (b)) In addition, the transfer-induced change in average income is independent of the transfer recipient (channel (c)).

For a size-weighted heterogeneous population with \( w_i = q_i \), all three channels are reflected in the first-order conditions. The “bottom fill-up”-rule therefore secures optimality only for the case of a homogeneous distribution. For heterogeneous distributions, the optimal transfer scheme can be derived using constrained optimization techniques. If \( I(\cdot) \) is convex, the planner’s problem can be solved with an interior-point algorithm. If \( I(\cdot) \) is quasiconvex, the bisection method can be used. Table 1 shows the categorization of several well-known inequality measures with respect to whether they are convex or quasiconvex.\(^7\)

\(^7\)For a needs-weighted heterogeneous population \( w_i = ES_i \forall i \), \( \frac{\partial u}{\partial t_i} = -\frac{2}{W+1} \).

\[
\frac{\partial v}{\partial t_i} = \frac{1}{W - W/N} \left[ \left( (F - \bar{F}) + \left( \frac{y_i + t_i}{ES_i} - \bar{y} \right) f \left( \frac{y_i + t_i}{ES_i} \right) \right) - \sum_{j=1}^{N} \frac{w_jw_i}{ES_iW} \left( (F - \bar{F}) + \left( \frac{y_j + t_j}{ES_j} - \bar{y} \right) f \left( \frac{y_j + t_j}{ES_j} \right) \right) \right]
\]

and \( \frac{\partial \bar{y}}{\partial t_i} = \frac{\partial \bar{y}}{\partial t_j} \forall i, j \). So, the redistributive effect of \( t_i \) depends on channels (a) and (b) but not on (c).

\(^8\)See Appendix A for proofs.
Table 1: Properties of Selected Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Convex</th>
<th>Quasiconvex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Absolute Mean Deviation</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Relative Mean Deviation</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Gini Coefficient</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Theil Index</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Atkinson Index</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

2.1. Solution for Convex Indices

Following Boyd and Vandenberghe (2004), an interior-point algorithm solves convex optimization problems of the type,

\[
\begin{align*}
\text{minimize} & \quad f \left( \left\{ x_n \right\}_{n=1}^M \right) \\
\text{subject to} & \quad g_j \left( \left\{ x_n \right\}_{n=1}^M \right) \leq 0, \quad j = 1, \ldots, J
\end{align*}
\]

where \( f(\cdot) \) and the functions \( g_j(\cdot) \) map from \( \mathbb{R}^M \) to \( \mathbb{R} \) and are twice continuously differentiable and convex.

In the context of the aforementioned social planner’s problem the program would be stated as,

\[
\begin{align*}
\text{minimize} & \quad I \left( \left\{ w_i, \frac{y_i + t_i}{ES_i} \right\}_{i=1}^N \right) \\
\text{subject to} & \quad 0 \leq t_i, \quad i = 1, \ldots, N \\
& \quad \sum_{i=1}^N t_i \leq B.
\end{align*}
\]

Ideally, we would proceed to optimize a modified Lagrangian with first-order conditions given by equation 4. If we knew the optimal values of the Lagrange multipliers, \( \nu_i^* \) and \( \lambda^* \), we could solve the first-order conditions for the optimal \( t_i^* \), since

\[
\frac{dI \left( \left\{ w_i, \frac{y_i + t_i^*}{ES_i} \right\}_{i=1}^N \right)}{dt_i} - \lambda^* + \nu_i^* = 0 \quad \forall i.
\]

However, we know neither \( \lambda^* \) nor \( \nu_i^* \), so we need a way to work around this issue. To deal with the inequality constraints, reformulate the objective function by adding terms that increase the value of the objective function when the \( t_i \) approach the boundary of the constraint set. This reformulation is the so-called logarithmic barrier and is one of the most straightforward implementations of an interior-point algorithm.\(^9\) The reformulated problem is,

\(^9\)This section in no way aims to supply the reader with a full set of instructions to implement the method. We want to foster understanding of why the method works and how it solves the problem of the planner.
minimize $\bar{I}(t) = I \left( \left\{ \frac{w_i, y_i + t_i}{ES_i} \right\}^N_{i=1} \right) + \iota \left( p, \{ t_i \}^N_{i=1} \right)$ \hspace{1cm} (10)

where $\iota \left( p, \{ t_i \}^N_{i=1} \right) = -\sum_{i=1}^{N} \frac{1}{p} \log(t_i) - \frac{1}{p} \log \left( B - \sum_{i=1}^{N} t_i \right)$ is a reasonable approximation to the indicator function of the $t_i$ being inside the set of constraints, depending on how large $p > 0$ is. $\iota(\cdot)$ will punish a step toward the boundary of the constraint set by increasing the function value.\textsuperscript{11}

The connection between equation (10) and the modified Lagrangian becomes transparent if the objective function is restated. An equivalent problem to equation (10) is $p I (\cdot) - \sum_{i=1}^{N} \log(t_i) - \log \left( B - \sum_{i=1}^{N} t_i \right)$. Now assume that vector $\{t_i\}^N_{i=1}$ solves the optimization problem and that $\{t_i\}^N_{i=1}$ is within the constraint set, then the following holds,

$$
\frac{dI \left( \left\{ \frac{w_i, y_i + t_i}{ES_i} \right\}^N_{i=1} \right)}{dt_i} - \frac{1}{t_i^*} - \left( \frac{-1}{B - \sum_{i=1}^{N} t_i^*} \right) = 0 \forall i.
$$

To see how the first-order condition (11) resembles the first-order conditions (4), set $\nu_i^* = \frac{1}{pt_i^*}$ and $\lambda^* = \frac{-1}{p(B - \sum_{i=1}^{N} t_i^*)}$. Then,

$$
\frac{dI \left( \left\{ \frac{w_i, y_i + t_i}{ES_i} \right\}^N_{i=1} \right)}{dt_i} - \lambda^* + \nu_i^* = 0 \forall i,
$$

which is the necessary and sufficient condition for the optimum of the modified Lagrangian.

Problem (10) can be solved by Newton’s method, proceeding through the three steps detailed below:

**Step 1** Start from a feasible transfer schedule, $t_0$, for example $t_i = B/N \forall i$. This gives an initial value of $I(\cdot)$, $I_0$. To improve $I_0$, calculate the Newton step, defined as,

$$
\Delta t = - \left( \nabla^2 I(t) \right)^{-1} \nabla I(t).
$$

The Newton step is the product of two terms, the inverse of the Hessian, $\left( \nabla^2 I(t) \right)^{-1}$, and the negative gradient, $-\nabla I(t)$. $\Delta t$ gives the change of the transfer schedule $t_0$ (to be multiplied with step-size scalar, $s$). Thus the Newton step gives the direction toward the optimum, defined by the weighted negative gradient. Weighting by the Hessian provides the benefit of fitting the steps to the shape of the contour sets of the function, as shown in Figure 1, thereby inducing fast convergence. The ellipse around $t_0$ is given by the points of unit distance measured in terms of the norm of the Hessian. Step 1 selects the point on the boundary of the ellipse that gives the smallest value of the objective function.

**Step 2** The line through the initial point $t_0$ and the point on the boundary of the ellipse is defined by $t_0 + s\Delta t$ (black line in Figure 1). Step 2 searches for the optimal $s$, $s^*$, that determines the point on the line

\textsuperscript{11}For the method to operate correctly, we require starting values of the $t_i$ to lie in the constraint set.
with the smallest level of $I$. This is the so-called line-search. One line-search variant is called backtracking. It relies on the idea of approximating $I$ along the direction $\Delta t$ with a first-order Taylor expansion. Set $s = 1$ and compute $\hat{I}$. If the transfer-induced reduction of the objective function is sufficiently large, the line-search algorithm terminates. Sufficiently large means that the new value of the objective function is smaller than $\alpha \in (0, 1)$ is set by the researcher to define the acceptable decrease of $\hat{I}$ along the current direction $\Delta t$. Otherwise, rescale $s$ with parameter $\beta \in (0, 1)$, and repeat the above procedure. Again, $\beta$ is determined by the researcher. Completing step 2 gives the new point $t_1 = t_0 + s^* \Delta t$.

**Step 3** Here the decrease of $\hat{I}$ between transfer scheme $t_0$ and $t_1$ in terms of the squared Newton decrement, defined as $ND^2(t) = \nabla^2 \hat{I}(t)$, is evaluated. If the squared Newton decrement is smaller than the specific threshold, $2\epsilon$, with $\epsilon$ defined by the researcher, then the algorithm terminates. Otherwise, start with $t_1$ and repeat steps 1-3 again.

Newton’s method with backtracking line search is summarized in the box below.

**Algorithm 1** Newton’s Method with Backtracking Line-Search

Start from some point $t$ in the domain of $I(\cdot)$ and choose $\epsilon > 0$ as the tolerance.

```
Compute $\Delta t$ and $ND^2(t)$ at the current point.
while $ND^2(t)/2 > \epsilon$ do
  1. Line-Search: Choose some $\alpha \in (0, 1)$ and a $\beta \in (0, 1)$ Set $s = 1$.
     while $\hat{I}(t + s \Delta t) > \hat{I}(t) + \alpha s \nabla \hat{I}(t)' \Delta t$ do
        Set $s = \beta s$.
     end while
  2. Update: Change the current point $t$ to be $t = t + s \Delta t$
end while
```

### 2.2. Solution for Quasiconvex Indices

Solutions to quasiconvex optimization problems exploit the structure of the quasiconvex function, turning the problem into a sequence of convex optimization problems: since every quasiconvex function has convex sublevel sets, there exists a convex function $\Phi_k(I(t), k)$ for some fixed $k$. For example, the Gini index is quasiconvex, but $\sum_{j=1}^N w_j \sum_{i=1}^N w_i \left| \frac{w_i + t_j}{ES_i} - \frac{w_i + t_j}{ES_j} \right| - k2W \sum_{i=1}^N w_i \frac{w_i + t_j}{ES_i}$ is convex. Finding the lowest $k$ in the sequence of convex optimization problems defined by $k$ is the objective. This is achieved through the so-called bisection method, which proceeds as follows:

**Step 1** Determine an upper and lower bound of the inequality index, denoted $u$ and $l$. For example, in the case of the Gini, one can set $l = 0$ and $u$ to its initial pre-transfer value. Further, define a level of tolerance as stopping criterion, $\epsilon > 0$.

---

\[^{12}\alpha\text{ cannot be set to be larger than 1, since, at best, the condition } \hat{I}(t + s \Delta t) = \hat{I}(t) + s \nabla \hat{I}(t)' \Delta t \text{ can be fulfilled with an infinitesimal } s. \text{ The Taylor expansion is a lower bound on the function } \hat{I}.\]

\[^{13}\text{Backtracking is quite insensitive to the choice of } \alpha \text{ and } \beta. \text{ See Boyd and Vandenberghe (2004).}\]

\[^{14}\text{See Appendix A.}\]
Figure 1: Newton’s Method with Backtracking Line-Search for the Function $f(t) = e^{t_1+3t_2-0.1} + e^{t_1-3t_2-0.1} + e^{-t_1-0.1}$. Dotted lines are contour sets of $f(t)$. The ellipse around $t_0$ is given by $\{x \mid \|x - t_0\| \|\nabla^2 f(t_0)\| = 1\}$. The black line is given by $t_0 + s\Delta t$.

**Step 2** Compute the candidate value for the sublevel set, $k = \frac{l+u}{2}$.

**Step 3** Use an interior-point algorithm to solve the inequality-constrained optimization problem: minimize $0$ s.t. $\Phi_k(t) \leq 0$, $\sum_{i=1}^{N} t_i \leq B$, $t_i \geq 0 \forall i$. This is also called a feasibility problem, since any point will be optimal if it lies within the feasible set. If a feasible transfer scheme is found, set $u = k$. Otherwise, set $l = k$.

**Step 4** Repeat steps 2-3 until $u - l < \epsilon$.

The bisection method is detailed in the box below, and converges after $\log_2 (u - l/\epsilon)$ iterations.

**Algorithm 2** Bisection Method

Set the value $l$ as the lower bound of the function and $u$ as the upper bound. Define $\epsilon > 0$ to be the tolerance.

while $u - l \geq \epsilon$ do

1. Set $k = \frac{l+u}{2}$.

2. Minimize $0$ s.t. $\Phi_k(I(t)) \leq 0$, $\sum_{i=1}^{N} t_i \leq B$, $t_i \geq 0 \forall i$

   if the $x_n$ found in the last step satisfy the constraints then
   set $u = k$
   else
   set $l = k$

Below we give an illustration of the bisection method. The illustration assumes three iterations. In the first two iterations the feasibility problem is solved and new upper bounds $u_1$ and $u_2$ are obtained. In the third iteration no feasible solution is found, leaving the interval $[l_3, u_3]$, which satisfies the tolerance condition.

### 3. Application

This section presents two applications of constrained optimization techniques in the context of inequality. The first application relies on a synthetic dataset and serves two purposes: 1) to illustrate the difference
between a transfer-scheme based on bottom fill-up and the optimal transfer scheme; and 2) to give an impression of the optimization problem’s computational burden as a function of sample size. The second application is a real-world implementation.

The synthetic dataset is presented in first four columns of Table 2. The planner’s exogenous budget is 100 monetary units. Two indices serve as criteria: the quasiconvex Gini coefficient and the convex absolute mean deviation (AMD). The adjacent column provides the transfers under bottom fill-up, while the final column provides the optimal transfer scheme.

Table 2: Synthetic Data

<table>
<thead>
<tr>
<th>$y_i$</th>
<th>$q_i$</th>
<th>$ES_i$</th>
<th>$y_i/ES_i$</th>
<th>$t_{fill-up}$</th>
<th>$t_{opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>4</td>
<td>2</td>
<td>90</td>
<td>73.34</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>26.66</td>
<td>0</td>
</tr>
<tr>
<td>400</td>
<td>3</td>
<td>1.8</td>
<td>160</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>1</td>
<td>1</td>
<td>300</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>450</td>
<td>2</td>
<td>1.5</td>
<td>300</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>800</td>
<td>4</td>
<td>2</td>
<td>400</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>600</td>
<td>1</td>
<td>1</td>
<td>600</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1100</td>
<td>1</td>
<td>1</td>
<td>1100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Inequality Measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>159.86</th>
<th>158.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gini</td>
<td>0.3415</td>
<td>0.3381</td>
</tr>
</tbody>
</table>

To gain an impression of the computational burden of the optimization, we sequentially take n-folds of the original number of observations up to 2048 observations, while proportionally adjusting the transfer budget. Figure 3 shows the computational burden in seconds using a computer with an Intel i7-4771 (3.5GHz) processor, 8 GB RAM, Windows 10, and Matlab R2014b. It reveals that computer-time should not be an issue for the applicability of the optimization procedures detailed above.

The real-world application deals with the potential re-distributive effects of a German pension reform, the so-called Riester scheme, subject of the study in Corneo et al. (2015): For dependently employed individuals,
the scheme grants allowances and tax cuts based on saving efforts and their household’s composition. The policy was supposed to have a progressive effect on the German income distribution. Corneo et al. (2015) find that the progressive effect of the Riester scheme on the distribution of equivalent incomes is quite small: For 2010, they find a reduction of the Gini coefficient between the pre-Riester and post-Riester distribution from 0.32960 to 0.32946. The total transfer volume is 2,790 million €. The question arises if the small effect (0.04 percent reduction of the Gini coefficient) is due to the small transfer budget (compared to aggregate household income) or an noneffective allocation (in terms of inequality reduction). Applying the bisection method results in a Gini coefficient of 0.32633. Accordingly, the budget would have allowed an almost 1 percent Gini reduction.

4. Conclusion

One important aim of many transfer schemes is the reduction of inequality in the distribution of market incomes. Because populations are heterogeneous, meaning households differ in composition and needs, simply transferring income to the bottom of the distribution fails to guarantee a maximum inequality reduction. Here we have suggested and implemented procedures that identify the optimal transfer schemes in terms of inequality reduction relying on the interior-point algorithm and the bisection method, depending on whether the inequality index of interest is convex or quasiconvex. In application, we show that computer time should not undermine the applicability of the detailed procedures.
Appendix

A. Proofs of Convexity and Quasiconvexity

The following proofs show the property of convexity or quasi-convexity of a few inequality measures. We use $y_i$ to denote the income of an individual or a household. The proofs are without loss of generality as one could easily replace $y_i$ with $\frac{y_i + t_i}{E(S)}$ and scale the sums with weights $w_i$ without changing the outcomes of the proofs. Using $y_i$ simply makes the exposition less convoluted.\(^\text{16}\)

A.1. The Variance is Convex

The variance is defined as,

$$V\left(\{y_i\}_{i=1}^N\right) = \frac{1}{N} \sum_{i=1}^N \left( y_i - \frac{1}{N} \sum_{i=1}^N y_i \right)^2. \quad (14)$$

The functions $y_i - \frac{1}{N} \sum_{i=1}^N y_i$ are affine and therefore convex for all $i$. Squaring these functions and then summing preserves convexity. Therefore the variance is convex.

A.2. The Absolute Mean Deviation is Convex

The absolute mean deviation is defined as,

$$\text{AMD}\left(\{y_i\}_{i=1}^N\right) = \frac{1}{N} \sum_{i=1}^N \left| y_i - \frac{1}{N} \sum_{i=1}^N y_i \right|. \quad (15)$$

Now the functions $y_i - \frac{1}{N} \sum_{i=1}^N y_i$ are composed with the absolute value, which is a norm. Norms are convex and, therefore, the convexity of the expression is preserved.\(^\text{17}\) As before with the variance, this implies that the absolute mean deviation as a whole is convex.

A.3. The Gini Coefficient is Quasiconvex

One definition of the Gini Coefficient is

$$G\left(\{y_i\}_{i=1}^N\right) = \frac{1}{2N \sum_{i=1}^N y_i} \sum_{i=1}^N \sum_{j=1}^N |y_i - y_j|. \quad (16)$$

To establish quasiconvexity of $G(\cdot)$, we need to establish that $G(\cdot)$ is quasiconvex in the $y_i$. To make progress we need to introduce sub-level sets of $G(\cdot)$. Call these $\Gamma$.

$$\Gamma = \{ x \in S : G(x) \leq \gamma \}, \quad S = \text{dom}(G(\cdot)) \quad (17)$$

---

\(^{16}\)If we optimize with respect to the $t_i$, then $\frac{y_i + t_i}{E(S)}$ is just an affine transformation of the $t_i$ and therefore preserves concavity or convexity. Changing the sums to be weighted also preserves concavity or convexity. See Boyd and Vandenberghe (2004).

\(^{17}\)See Boyd and Vandenberghe (2004).
If the elements of $\Gamma$ are convex for every $\gamma$, then $G(\cdot)$ is quasiconvex. The sub-level set for an arbitrary $\gamma$ may be denoted by

$$\frac{1}{2N \sum_{i=1}^{N} y_i} \sum_{j=1}^{N} \sum_{i=1}^{N} |y_i - y_j| \leq \gamma \quad (18)$$

$$\sum_{j=1}^{N} \sum_{i=1}^{N} |y_i - y_j| - \gamma 2N \sum_{i=1}^{N} y_i \leq 0 \quad (19)$$

This holds since the mean is non-negative. This expression can be shown to describe a convex set in $y_i$ by establishing that the left-hand side is a convex function in $y_i$ for all $\gamma$. This is sufficient, since any sub-level set of a convex function is a convex set and here we are studying the sub-level set of the function with level-value zero.

To advance, we need to re-express the left-hand side as a function that has known convexity properties. We note that the left-hand side can be expressed as the point-wise maximum of $2^{N-1}$ linear expressions in $y_i$ with another linear function in $y_i$ subtracted. For example if $N = 2$:

$$\max \{2 (y_1 - y_2), 2 (y_1 - y_2)\} - \gamma 4 \sum_{i=1}^{2} y_i \quad (20)$$

The point-wise maximum of linear expressions is convex, so the maximum term is convex.$^{18}$ The second term is linear in the $y_i$ and therefore also convex. So the whole left-hand side is convex for any $\gamma$. Ergo, the Gini coefficient is quasiconvex in the $y_i$.

**A.4. The Relative Mean Deviation is Quasiconvex**

This is builds on the result that the absolute mean deviation is convex. Since $RMD = \frac{AMD(\{y_i\}_{i=1}^{N})}{\frac{1}{N} \sum_{i=1}^{N} y_i}$, we can form the sublevel sets

$$AMD(\{y_i\}_{i=1}^{N}) - \alpha \frac{1}{N} \sum_{i=1}^{N} y_i \leq 0, \quad (21)$$

where the lefthand side contains only convex terms making the sub-level sets of the RMD convex. Therefore, the RMD is quasiconvex.

**A.5. The Atkinson Index is Quasiconvex**

The Atkinson-Index is defined as

$$A_\epsilon(\{y_i\}_{i=1}^{N}) = 1 - \frac{1}{\frac{1}{N} \sum_{i=1}^{N} y_i} \left( \frac{1}{N} \sum_{i=1}^{N} (y_i)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}. \quad (22)$$

$^{18}$See Boyd and Vandenberghe (2004).
First, consider only the second term of $A_\epsilon$ and substitute $p = 1 - \epsilon$. Then

$$s \left( \{y_i\}_{i=1}^N \right) = \left( \frac{1}{N} \sum_{i=1}^{N} (y_i)^p \right)^{\frac{1}{p}}. \tag{23}$$

This function is concave whenever $(p - 1) < 0$ or equivalently $\epsilon \geq 0. \tag{19}$ To show quasiconvexity then, we need to establish that the sub-level sets of $A_\epsilon$ are convex. This is sufficiently shown by verifying that the negative term of $A_\epsilon$ has convex sub-level sets, as the rest is just an affine transformation. The sub-level sets are given by

$$- \frac{1}{N} \sum_{i=1}^{N} y_i \left( \frac{1}{N} \sum_{i=1}^{N} (y_i)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \leq \alpha \tag{24}$$

$$- \left( \frac{1}{N} \sum_{i=1}^{N} (y_i)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} - \alpha \frac{1}{N} \sum_{i=1}^{N} y_i \leq 0 \tag{25}$$

Now we determine whether the functions on the left-hand side are convex. If they are, they generate sets that are convex given any $\alpha$, which implies quasiconvexity. Since the first function is convex – the negative of $s \left( \{y_i\}_{i=1}^N \right)$ is convex – and the second function is affine, this is the case.

**A.6. The Theil Index is Quasiconvex**

The definition of the Theil Index is

$$T = \frac{1}{N} \sum_{i}^{N} \frac{N y_i}{\sum_{i}^{N} y_i} \log \left[ \frac{N y_i}{\sum_{i}^{N} y_i} \right]. \tag{26}$$

For quasiconvexity the sublevel sets of the Theil Index need to be convex. Accordingly,

$$\frac{1}{N} \sum_{i}^{N} \frac{N y_i}{\sum_{i}^{N} y_i} \log \left[ \frac{N y_i}{\sum_{i}^{N} y_i} \right] \leq \alpha \tag{27}$$

$$\sum_{i}^{N} y_i \log \left[ \frac{N y_i}{\sum_{i}^{N} y_i} \right] - \alpha \sum_{i}^{N} y_i \leq 0. \tag{28}$$

The functions on the left-hand side induce convex sets if they are convex. The second term is affine and therefore convex. The first term needs further investigation. It is convex if its Hessian is positive semidefinite. The second partial derivatives of $f(\{y_i\}_{i=1}^N) = \sum_{i}^{N} y_i \log \left[ \frac{N y_i}{\sum_{i}^{N} y_i} \right]$ are

$$f_{y_i,y_i} = \frac{1}{y_i} - \frac{1}{\sum_{i}^{N} y_i}, \quad f_{y_i,y_j} = -\frac{1}{\sum_{i}^{N} y_i}. \tag{29}$$

\[\text{See Boyd and Vandenberghe (2004, p. 87).} \]
Then the Hessian of \( f(\{y_i\}_{i=1}^N) \) is

\[
H_f = \begin{pmatrix}
\frac{1}{y_1} - \frac{1}{\sum_i^n y_i} & -\frac{1}{y_2} - \frac{1}{\sum_i^n y_i} & \cdots \\
\frac{1}{y_2} & \frac{1}{y_2} & \cdots \\
\vdots & \vdots & \ddots
\end{pmatrix}.
\] (30)

We can reshape the matrix before we test for positive semi-definiteness as

\[
H_f = \begin{pmatrix}
\frac{1}{y_1} & 0 & \cdots \\
0 & \frac{1}{y_2} & \cdots \\
\vdots & \vdots & \ddots
\end{pmatrix} - \begin{pmatrix}
\frac{1}{\sum_i^n y_i} & \frac{1}{\sum_i^n y_i} & \cdots \\
\frac{1}{\sum_i^n y_i} & \frac{1}{\sum_i^n y_i} & \cdots \\
\vdots & \vdots & \ddots
\end{pmatrix}.
\] (31)

The Hessian is positive semi-definite iff for any vector \( \mathbf{v} \),

\[
\mathbf{v}' \begin{pmatrix}
\frac{1}{y_1} & 0 & \cdots \\
0 & \frac{1}{y_2} & \cdots \\
\vdots & \vdots & \ddots
\end{pmatrix} \mathbf{v} - \mathbf{v}' \begin{pmatrix}
\frac{1}{\sum_i^n y_i} & \frac{1}{\sum_i^n y_i} & \cdots \\
\frac{1}{\sum_i^n y_i} & \frac{1}{\sum_i^n y_i} & \cdots \\
\vdots & \vdots & \ddots
\end{pmatrix} \mathbf{v} \geq 0.
\] (32)

To show this, we use the Cauchy-Schwarz-Inequality, which states that for any two vectors \( \mathbf{a} \) and \( \mathbf{b} \),

\[
(a' \mathbf{a})(b' \mathbf{b}) \geq (a' \mathbf{b})^2.
\] (33)

State the dot-product of the Hessian with \( \mathbf{v} \) as summations.

\[
\frac{1}{\sum_i^n y_i} \left( \left( \sum_i^N y_i \right) \left( \sum_i^N \frac{y_i^2}{y_i} \right) - \left( \sum_i^N y_i \right)^2 \right) \geq 0.
\] (34)

To complete the proof, pick \( \mathbf{a}' = (\sqrt{y_1}, \sqrt{y_1}, \ldots) \) and \( \mathbf{b}' = (\frac{y_1}{\sqrt{y_1}}, \frac{y_2}{\sqrt{y_2}}, \ldots) \), which establishes that the above sums are greater or equal to zero.

Since both functions determining the sublevel sets are convex for any \( \alpha \), the Theil is quasiconvex.

### B. Modification of Indices for Empirical Applications

Indices like the Gini coefficient in the following form,

\[
G = \frac{1}{2W \sum_{i=1}^N w_i} \sum_{i=1}^N w_i \sum_{j=1}^N w_j \left| \frac{y_i + t_i + ES_i}{ES_i} - \frac{y_j + t_j + ES_j}{ES_j} \right|.
\]

or the RMD and AMD involve the absolute value, which is not differentiable at zero. Since we require the objective functions to be twice continuously differentiable, we cannot simply proceed with such functions.

We address this issue by replacing the absolute value with a quadratic and then optimize; obtaining the optimal transfers for the original indices. This exploits the ordinal equivalence property of the absolute
value function and the quadratic function, which implies that the transfer scheme at the minimum of the modified index is the same for the original index.

The absolute value and the quadratic are ordinally equivalent iff $\forall a, b \in R$,

$$|a| \leq |b| \iff a^2 \leq b^2.$$  

This is true since the formula for the absolute value is

$$|\cdot| = \sqrt{\cdot^2}$$

and thus the quadratic is the result of applying a positive monotone transformation to the absolute value function, which preserves orderings.

The formula for the modified Gini coefficient thus is

$$G = \frac{1}{2W} \sum_{i=1}^{N} w_i \left( \sum_{j=1}^{N} w_j \left( \frac{y_i + t_i}{ES_i} - \frac{y_j + t_j}{ES_j} \right)^2 \right).$$

This modification also retains the (quasi)convexity properties of the indices, since the quadratic function is a convex function.

C. Matlab Code

The bisection method is implemented using three functions in Matlab. The central file is `Bisec.m`.

```matlab
function [ minGini ] = Bisec(y,es,w,u,l,e,A,b,st,LB)
while u-l > e
    k=(u+l)/2;
    x=fmincon(@(x)0,st,A,b,[],[],LB,[],@(x)cons(y,w,es,x,k),optimoptions('fmincon','Algorithm','interior-point','TolCon',10^(-7),'TolX',10^(-7)))
    cand=MGC(y,w,es,x);
    if cand <= k
        u=k
    else
        l=k
    end
end
minGini=[u,x];
end
```

One passes the data in terms of the variables $y$, $es$ and $w$. $u$ and $l$ are the upper and lower bounds for the bisection method. $e$ is the level of tolerance. $A$ and $b$ let us define the budget inequality constraint. $st$ is the vector of starting values and $LB$ is the lower bound for our variables.

The inequality constraints are implemented via the function `cons`. Here the sublevel set of the modified Gini coefficient (Appendix B) is calculated.
function [ c,ceq ] = cons( y,w,es,x,k )
S=2*sum(w)*(sum(w.*(es.^-1.*(y+x))));
yt=es.^-1.*(y+x);
S=0;
for i = 1:length(y)
    S=S+(w(i)*transpose(w)*(yt(i)-yt).^2);
end

c=(S)-k*s;
ceq=[];
end

Bisec checks the value of the modified Gini coefficient against the reference value k using the function MGC, which simply calculates the modified Gini coefficient.

function [ g ] = MGC(y,w,es,t)
S=2*sum(w)*(sum(w.*(es.^-1.*(y+x))));
yt=es.^-1.*(y+x);
S=0;
for i = 1:length(y)
    S=S+(w(i)*transpose(w)*(yt(i)-yt).^2);
end
g=S/s;
end
References


<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
<th>Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016/1</td>
<td>BARTELS, Charlotte und Maximilian STOCKHAUSEN</td>
<td>Children’s opportunities in Germany – An application using multidimensional measures</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/2</td>
<td>BÖNKE, Timm; Daniel KEMPTNER und Holger LÜTHEN</td>
<td>Effectiveness of early retirement disincentives: individual welfare, distributional and fiscal implications</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/3</td>
<td>NEIDHÖFER, Guido</td>
<td>Intergenerational Mobility and the Rise and Fall of Inequality: Lessons from Latin America</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/4</td>
<td>TIEFENSEE, Anita und Christian WESTERMEIER</td>
<td>Intergenerational transfers and wealth in the Euro-area: The relevance of inheritances and gifts in absolute and relative terms</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/5</td>
<td>BALDERMANN, Claudia; Nicola SALVATI und Timo SCHMID</td>
<td>Robust small area estimation under spatial non-stationarity</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/6</td>
<td>GÖRLITZ, Katja und Marcus TAMM</td>
<td>Information, financial aid and training participation: Evidence from a randomized field experiment</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/7</td>
<td>JÄGER, Jannik und Theocharis GRIGORIADIS</td>
<td>Soft Budget Constraints, European Central Banking and the Financial Crisis</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/8</td>
<td>SCHREIBER, Sven und Miriam BEBLO</td>
<td>Leisure and Housing Consumption after Retirement: New Evidence on the Life-Cycle Hypothesis</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/9</td>
<td>SCHMID, Timo; Fabian BRUCKSCHEN; Nicola SALVATI und Till ZBIRANSKI</td>
<td>Constructing socio-demographic indicators for National Statistical Institutes using mobile phone data: estimating literacy rates in Senegal</td>
<td>Economics</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
<td>Title</td>
<td>Journal</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>2016/10</td>
<td>JESSEN, Robin; ROSTAM-AFSCHAR, Davud und Sebastian SCHMITZ</td>
<td>How Important is Precautionary Labor Supply?</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/11</td>
<td>BIER, Solveig; Martin GERSCH, Lauri WESSEL, Robert TOLKSDORF und Nina KNOLL</td>
<td>Elektronische Forschungsplattformen (EFP) für Verbundprojekte: Bedarfs-, Angebots- und Erfahrungsanalyse</td>
<td>Wirtschaftsinformatik</td>
</tr>
<tr>
<td>2016/12</td>
<td>WEIDENHAMMER, Beate; Timo SCHMID, Nicola SALVATI und Nikos TZAVIDIS</td>
<td>A Unit-level Quantile Nested Error Regression Model for Domain Prediction with Continuous and Discrete Outcomes</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/13</td>
<td>TZAVIDIS, Nikos; Li-Chun ZHANG, Angela LUNA HERNANDEZ, Timo SCHMID, Natalia ROJAS-PERILLA</td>
<td>From start to finish: a framework for the production of small area official statistics</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/14</td>
<td>GASTEIGER, Emanuel</td>
<td>Do Heterogeneous Expectations Constitute a Challenge for Policy Interaction?</td>
<td>Economics</td>
</tr>
<tr>
<td>2016/15</td>
<td>HETSCHKO, Clemens; Ronnie SCHÖB und Tobias WOLF</td>
<td>Income Support, (Un-)Employment and Well-Being</td>
<td>Economics</td>
</tr>
</tbody>
</table>