Problem Set 1

1. Lorenz Curve and Gini Coefficient

a) See Figure at the end.

b) The results using the linear interpolation are:

<table>
<thead>
<tr>
<th>Year</th>
<th>Top 10 percent share</th>
<th>Top 1 percent share</th>
<th>90/50 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>42%</td>
<td>16%</td>
<td>3.31</td>
</tr>
<tr>
<td>2016</td>
<td>44%</td>
<td>19%</td>
<td>4.19</td>
</tr>
</tbody>
</table>

c) The results using the Pareto interpolation are:

<table>
<thead>
<tr>
<th>Year</th>
<th>Top 10 percent share</th>
<th>Top 1 percent share</th>
<th>90/50 ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>42%</td>
<td>17%</td>
<td>3.18</td>
</tr>
<tr>
<td>2016</td>
<td>47%</td>
<td>20%</td>
<td>3.30</td>
</tr>
</tbody>
</table>

Elaborate explanation on Pareto interpolations

We will use a Pareto interpolation following the method discussed in section 3 in Atkinson et al. (2011) and section 5.1 in Atkinson (2005). We will begin by describing the Pareto distribution and some of its useful properties, and then discuss the Pareto interpolation procedure.

The Pareto cumulative distribution function $F(y)$ for $y$ is,

$$F(y) = 1 - \left( \frac{k}{y} \right)^\alpha, \quad \alpha > 1, \ k > 0$$

$$\Rightarrow f(y) = \frac{\alpha k^\alpha}{y^{1+\alpha}}$$

where $k$ and $\alpha$ are given parameters, $\alpha$ is referred to as the Pareto parameter. The $\alpha$ parameter determines the thickness of the tail of the distribution.

The key property of the Pareto distribution is that the ratio of average income $y^*(y)$ of individ-
uals above \( y \) to \( y \) does not depend on the income threshold of \( y \) (\( k \))

\[
y^*(y) = \frac{\int_{z>y} zf(z)dz}{\int_{z>y} f(z)dz} = \frac{\int_{z>y} z\alpha k^\alpha / z^{1+\alpha}dz}{\int_{z>y} \alpha k^\alpha / z^{1+\alpha}dz} = \frac{\int_{z>y} 1/z^\alpha dz}{\int_{z>y} 1/z^{1+\alpha}dz}
\]

\[
= \frac{\frac{1}{1 - \frac{1}{\alpha z^\alpha}}\big|_y^\infty}{\frac{1}{1 - \frac{1}{\alpha z^\alpha}}\big|_y^\infty} = \frac{0 - \frac{1}{(\alpha-1)y^{\alpha-1}}}{0 - \frac{1}{ay^a}} = y \cdot \alpha / (\alpha - 1)
\]

\[
\Rightarrow \frac{y^*(y)}{y} = \beta, \text{ with } \beta = \alpha / (\alpha - 1)
\]

The parameter \( \beta \) has a more intuitive interpretation than \( \alpha \), if \( \beta = 2 \) the average income of individuals with income above $100,000 is $200,000. A higher \( \beta \) means a flatter upper tail of the income distribution. We will use \cite{Atkinson} notation. Denote by \( H(y) \) the share of individual with income higher than \( y \),

\[
H(y) = 1 - F(y) = \left(\frac{k}{y}\right)^\alpha = Ay^{-\alpha}, \text{ with } A = k^\alpha
\]

Denote by \( G(y) \) the total income of received by individuals with income \( y \) or higher. According to the properties of the Pareto distribution,

\[
G(y) = \frac{\alpha}{\alpha - 1} Ay^{-(\alpha - 1)}
\]

The share in total income of people with incomes \( y \) or higher can be obtained by dividing \( G(y) \) by \( \mu \), where \( \mu \) is the mean income, denote by \( \Omega(y) = \frac{G(y)}{\mu} \).

Let \( y_p \) the desired income percentile. There are to kinds of interpolations:

1. When we observe tabulations of incomes shares lower and higher then \( y_p \), in this case we can use interpolation between the two adjacent tabulations that \( y_p \) is in the middle of them. For example we observe the total income of the top 3.677% and the top 0.7688%, however we do not observe the total income of the top 1%.

2. The second scenario is when we want to estimate the share in total income of a percentile we observe only percentiles that are higher than it. For example the smallest percentile we observe is 0.0123%, hence if we want to estimate the share in total income of the top 0.001% we will have only tabulations from one side. This is referred as open range tabulation. In this exercise all the tabulations we are asked to perform are from the first type.

\footnote{Using the notation from \cite{Atkinson et al.}}
Pareto interpolations

I will begin by interpolating the share in total income of the top 1%. Denote by \( G_1, G_2, \) and \( G^* \) the share in total of the top 0.768%, 3.677% and 1% respectively. Denote by \( H_1 \) and \( H_2 \) the \( H(y) \) of the two tabulations to the left and right of \( H(y) = 1 \) (e.g., \( H_1 = 0.00768 \) and \( H_2 = 0.03677 \)).

It follows from the Parto distribution that

\[
\frac{\alpha}{\alpha - 1} = \log \left( \frac{H_1 / H_2}{G_1 / G_2} \right)
\]

Therefore using equation (2) we can estimate \( \frac{\alpha}{\alpha - 1} \) for the top 1 percent. Note, that this is making a Pareto interpolation between the two tabulations closest to the desired quantity.

\[
\beta = \log \left( \frac{H_1 / H_2}{G_1 / G_2} \right)
\]

\( \Rightarrow \alpha = \frac{\beta}{\beta - 1} \)

Next we use \( G_1 \) and \( H_1 \) to approximate the total earnings of the top 1% (it does not matter if we choose \( H_2 \) and \( G_2 \), the result would have been the same).

\[
\beta \approx \log \left( \frac{H_1 / 0.01}{G_2 / G^*} \right) \Rightarrow G^* = \log \left( \frac{G_2 / G^*}{G_1 / G_2} \right) = \frac{1}{\beta} \cdot \log \left( \frac{H_1 / 0.01}{G_2 / G^*} \right)
\]

\[
\Rightarrow G_2 / G^* = \exp \left( \frac{1}{\beta} \cdot \log \left( \frac{H_1 / 0.01}{G_2 / G^*} \right) \right) = (H_1 / 0.01)^{\frac{1}{\beta}}
\]

\[
G^* = G_2 \cdot (H_1 / 0.01)^{\frac{1}{\beta}}
\]

Next we use the same procedure as before to estimate \( \alpha \) for each percentile, and then use the CDF of the Pareto distribution. Denote by \( p \) the percentile of interest and by \( y_p \), income of the individual at the \( p \) percentile. Given a tabulation we can calculate \( \alpha \). Next we extrapolate the parameter \( k \) (or \( A \) in Atkinson’s notation) for each of the tabulations above and below the desired \( p \),

\[
k = (1 - F(y))^{1/\alpha} \cdot y
\]

---

\( ^2 \)Proof: Let \( y_1 \) and \( y_2 \) be two income levels such that \( y_1 > y_2 \) and therefore \( H(y_1) < H(y_2) \).

\[
\log \left( \frac{H_1 / H_2}{G_1 / G_2} \right) = \log \left( \frac{\frac{y_1 \alpha^\alpha}{y_2^\alpha}}{\frac{y_1 \alpha^\alpha}{y_2^\alpha}} \right) = \log \left( \frac{y_1 / y_2^{-\alpha}}{y_1 / y_2^{-\alpha}} \right) = \frac{\alpha \log \left( \frac{y_1 / y_2}{y_1 / y_2} \right)}{-(\alpha - 1) \log \left( \frac{y_1 / y_2}{y_1 / y_2} \right)} = \frac{\alpha}{\alpha - 1}
\]

---

3
After we have $k$ and $\alpha$ the desired income threshold, $y_p$ is,

$$y_p = \frac{k}{(1 - F(y_p))^{1/\alpha}}$$  \hspace{1cm} (3)$$

We need to decide which of the tabulations to use in order to calculate $k$. Piketty 2001, Appendix B, (Page 598 in the book) suggest to choose the tabulation that is closest to the desired percentile. It is clear that the choice of which tabulation to use has a significant impact on the results.

d) See second figure at the end of the solutions for the Lorenz curve.

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10 percent share</td>
<td>39%</td>
<td>41%</td>
</tr>
<tr>
<td>Top 1 percent share</td>
<td>14%</td>
<td>17%</td>
</tr>
<tr>
<td>90/50 ratio</td>
<td>3.31</td>
<td>4.19</td>
</tr>
</tbody>
</table>

e) Mis-ranking always under-estimates the Gini coefficient (as the Lorenz curve is furthest away from the diagonal precisely when individuals are perfectly ranked by income).

### 3. Optimal Top Income Tax Rate with Income Effects

This exercise follows Saez Restud’01 section 3 (see the paper for more details).

a) In the tax reform graph drawn in class, increasing $\tau$ by $d\tau$ creates a negative substitution effect (less slope) leading to less work, and a negative income effect leading to more work. Hence, $e$ is a mix of substitution and income effects.

b) Let $z^i(1 - \tau, R)$ be the earnings supply function obtained from solving the individual utility maximization problem under a linear tax:

$$\max_{c,z} u_i(c, z) \text{ st } c = (1 - \tau)z + R$$

We denote by $e^i_u$ the uncompensated elasticity of $z^i$ with respect to $1 - \tau$ and by $\eta^i = (1 - \tau)\partial z^i / \partial R$ the income effect parameter. As a simple graph shows (see Saez Restud’01 Section 3), the reform changes $1 - \tau$ by $-d\tau$ and changes $R$ by $dR = z^*d\tau$. Note, that the virtual income is defined as $R \equiv z^*\tau$ (in Saez 2001 it is written as $\bar{R} = \bar{z}\tau$), and as we assume that $z^*$ is fixed it follows immediately that $dR = z^*d\tau$. Hence, we have:

$$dz^i = -d\tau \frac{\partial z^i}{\partial(1 - \tau)} + z^*d\tau \frac{\partial z^i}{\partial R} = -\frac{d\tau}{1 - \tau} z^i \frac{1 - \tau}{z^i} \frac{\partial z^i}{\partial(1 - \tau)} + \frac{d\tau}{1 - \tau} z^*(1 - \tau) \frac{\partial z^i}{\partial R} = -\frac{d\tau}{1 - \tau} z^i e^i_u + \frac{d\tau}{1 - \tau} z^* \eta^i$$

c) Recall that the elasticity $e$ is defined as

$$e = \frac{1 - \tau}{\sum z^i d(1 - \tau)}.$$
Hence we have

\[ e = \frac{1 - \tau}{\sum z^i} \sum \left[ \frac{1}{1 - \tau} z^i e^i - \frac{1}{1 - \tau} z^i \eta^i \right] = \frac{\sum z^i e^i - z^i \sum \eta^i}{\sum z^i} \]

with \( N \) number of top bracket taxpayers. Hence \( e = \bar{e}_u - \frac{a - 1}{a} \bar{\eta} \) with \( \bar{e}_u \) the income weighted average of \( e^i \) and \( \bar{\eta} \) the straight average of \( \eta^i \) among top bracket taxpayers.

The uncompensated elasticity is income weighted because those with higher income should count more in the response. In contrast, the income effect parameter is not an elasticity and hence should not be income weighted.

d)-e) Suppose, each individual utility takes the form \( u^i(c, z) = \log(c) - a^i \cdot z^{1+1/\varepsilon} \) with \( a^i \) is a parameter (that can vary across individuals). The individual solves

\[ \max_z u^i((1 - \tau)z + R, z) = \log((1 - \tau)z + R) - a^i \cdot z^{1+1/\varepsilon} \]

The individual FOC in \( z \) is:

\[ \frac{1 - \tau}{(1 - \tau)z + R} = a^i(1 + 1/\varepsilon)z^{\frac{1}{\varepsilon}} \]

or

\[ \log(1 - \tau) - \log((1 - \tau)z + R) = \frac{1}{\varepsilon} \log(z) + \log[a^i(1 + 1/\varepsilon)] \]

This defines implicitly \( z^i(1 - \tau, R) \) with the following comparative statics. A small change \( dR \) leads to \( dz^i \) such that:

\[ dz^i \left[ \frac{1}{\varepsilon z} + \frac{1 - \tau}{(1 - \tau)z + R} \right] = -\frac{dR}{(1 - \tau)z + R} \]

Hence

\[ \eta^i = (1 - \tau) \frac{dz^i}{dR} = -\frac{1}{1 + \frac{z + R/(1 - \tau)}{\varepsilon z}} \rightarrow z \rightarrow \infty \frac{\varepsilon}{\varepsilon + 1} \]

A small change \( d(1 - \tau) \) leads to \( dz^i \) such that:

\[ dz^i \left[ \frac{1}{\varepsilon z} + \frac{1 - \tau}{(1 - \tau)z + R} \right] = \frac{d(1 - \tau)}{1 - \tau} - \frac{zd(1 - \tau)}{(1 - \tau)z + R} \]

Hence

\[ (1 - \tau) \frac{dz^i}{d(1 - \tau)} \left[ \frac{1}{\varepsilon z} + \frac{1}{z + R/(1 - \tau)} \right] = 1 - \frac{z}{z + R/(1 - \tau)} \]

\[ e^i_u = \frac{1 - \tau}{z} \frac{dz^i}{d(1 - \tau)} = -\frac{1 - \frac{z}{z + R/(1 - \tau)}}{\varepsilon + \frac{z}{z + R/(1 - \tau)}} \rightarrow z \rightarrow \infty 0 \]

e) Use c) to obtain an optimal top tax rate formula as a function of \( \varepsilon \) and the Pareto parameter \( a \) in that case. We have \( e = \bar{e}_u - \frac{a - 1}{a} \bar{\eta} \approx \frac{a - 1}{a} \frac{\varepsilon}{\varepsilon + 1} \) and hence:

\[ \tau = \frac{1}{1 + ae} = \frac{1}{1 + (a - 1)\frac{\varepsilon}{\varepsilon + 1}} = \frac{\varepsilon + 1}{1 + a\varepsilon} \]
If $a = 1.5$ and $\varepsilon = 1$, we have $\tau = 1/(1 + .5 \cdot .5) = 80\%$.

With $\varepsilon = \infty$, we get $\tau = 1/(1 + .5) = 66.6\%$.

This utility specification has zero uncompensated elasticity and hence large income effects when the compensated elasticity is large. As a result, the tax rate is substantial, even with a very large Frisch elasticity.
Figures

2016 Data

Gini coefficient: 0.5882

Cumulative share of returns from lowest to highest AGI (Only positive AGI included)

2003 Data

Gini coefficient: 0.5555

Cumulative share of returns from lowest to highest AGI (Only positive AGI included)

2016 Data – After tax

Gini coefficient: 0.5548

Cumulative share of returns from lowest to highest AGI (Only positive AGI included)

2003 Data – After tax

Gini coefficient: 0.5261

Cumulative share of returns from lowest to highest AGI (Only positive AGI included)
References
