

Mahdi Soleimani - Homework 04

Part 1: Explanations

1 Sampling and outliers (3 points)

(a) A few problems arise. First of all, it is difficult to be conclusive when analyzing over time. Since the same households are not followed over time, we cannot make sure whether changes in the variable are systematic or random. Secondly, some variables by definition change over time. If we do not follow the same household, it becomes difficult to fully understand the impact of the variable. For instance, age influences many behaviors. However, in the repeated cross-section data, we cannot see how one young individual changed his behavior (for instance, savings) over time as he got older. Thirdly, we cannot fully determine the extent to which exogenous shocks (civil war, earthquake, etc) influence variables. If we follow the same household before and after a shock, then it becomes much easier to understand their relationship. (For more information, see <https://www.ukdataservice.ac.uk/media/455362/changeovertime.pdf>)

(b) Outliers on both sides, that is, the very poor and the very rich should be omitted from the data set. First of all, there are measurement issues with the outliers. For instance, it is hard to capture all the sources of income or the entire amount of consumption. Also, the rich tend to understate and the poor tend to overstate, maybe because they want to show themselves closer to average. Secondly, most inequality measures are sensitive to outliers. Therefore, if the mean of the population has to be estimated from the sample, we could have problems. In general, inequality increases as measurement errors increase.¹

(c) In order to obtain consumption per household, we divide total consumption ("constot" in the data set) by household size ("hhsz").

(d) The distribution of income is skewed to the right. For instance see the graphs in 2008. Taking the log of consumption per capita will increase the space between variables that are low and decrease the space between variables that are high. As a result, it will change the distribution into normal (approximately, see page 4). Therefore if we exclude values that are 3σ away from the median, given the properties of the normal distribution, we would end up excluding only 1% of the data. Therefore, we will not lose many observations. This is a statistically robust correction. 121 observations were excluded. The reason we do not use the weights is that we want to exclude the outliers from the sample. The weights, however, are related to the population.

2 Consumer Price Index and Purchasing Power Parities (4 points)

(a) A representative household has 4 members, is male-headed and whose head is a blue-collar worker; either self-employed outside agriculture or a farmer.

In order to calculate the weighted average of the share of food expenditure, we do the following calculations. The first fraction is the definition (w is the weight of the household). However, we can use the per capita version of the variables, simply by multiplying by n . Also, we can multiply and divide household consumption per capita in the numerator without changing anything. Now we have our desired weight: $W = w \cdot n \cdot \text{household consumption per capita}$.

$$= \frac{\sum_{hh} w \cdot \text{food expenditure}}{\sum_{hh} w \cdot \text{household consumption}}$$
$$= \frac{\sum_{hh} w \cdot n \cdot \text{food expenditure per capita}}{\sum_{hh} w \cdot n \cdot \text{household consumption per capita}}$$

¹This fact is shown in Cowell, F. and Victoria-Feser, M.P., 1993. Robustness properties of inequality measures: The influence function and the principle of transfers.

$$\begin{aligned}
&= \frac{\sum_{hh} w.n.household \text{ consumption per capita. } (\frac{food \text{ expenditure per capita}}{household \text{ consumption per capita}})}{\sum_{hh} w.n.household \text{ consumption per capita}} \\
&= \frac{\sum_{hh} W. \text{ share of food expenditure}}{\sum_{hh} W}
\end{aligned}$$

(b) and (c)

When calculating the CPI, the Paasche index often understates the inflation whereas the Laspeyres index often exaggerates. As prices increase over time, people tend to adjust their baskets, i.e, they decrease the quantity of the goods that went through a price increase. If we do not change the basket, as in Laspeyres, we end up with a higher index. On the other hand, since we only use quantity of time t in Paasche, we end up with a lower index. In our data, this true in 1998. Also, the CPI in the South is always greater than in the North.

With regard to PPP, we have different results for the North and and the South because of price differences. This is because in different regions, people develop different preferences and different resources are available. As a result, calculating the prices for a given basket in each region will result in different indices. The Fisher index is the geometric mean of Paasche and Laspeyres. Therefore, it is always between the two.

(d) We must use the following equation.

$$conspc87 = consumption \text{ per capita. } \frac{PPP}{CPI}$$

Dividing the consumption per capita by CPI allows us to compare numbers through the years. Since the purchasing power of the North and the South differ, we have to multiply the result by PPP. For instance, if 100 CFA franc is worth (in terms of buying goods) 73 in the North, X CFA franc is worth 0.73X in the North. This allows us to compare the regions. Implementing them both will harmonize the data in PPP 1987 franc CFA.

3 Absolute poverty measures (5 points)

(a) and (b)

Here, we do not need to convert 273 by using CPI since it is in 1987 franc CFA. However, we need to adjust for PPP. Therefore, we need to have two poverty lines: 237 for the South and $2.73 \times 72.91 = 172.7967$ for the North. We do similar operations for 474 to get the two poverty lines. We expect if we move the poverty line higher, we end up with a higher poverty rate. Our computation proves this guess.

Whether we take 1\$ or 2\$ per day as the threshold, the evolution is similar. At the national level, in the South, and in the North, the poverty is increasing. In particular, the increase in the 1987-1998 has been dramatic. North is always poorer than the south.

(c)

This measure also captures the evolution of poverty we talked about. Comparing the results to part (b), we see that in 1987 for each region, the poverty rates are the same, we are using the same CPI (100 for both north and south), and PPP is 100. Also, the only factor that determines differences poverty in each region is consumption per capita.

For the years 1998 and 2008, PPP remains 100 but we use the CPI of the south for both regions to account for inflation. For the south, poverty rates are similar because the CPI we used is the same as part a. However, for the North as well as the national level, they change because we are using the CPI of the south. As a result, if the original CPI is larger than the the CPI of the south, poverty rates increase when we use the CPI of the south. However the general trend we explained above holds here as well.

4 Growth Incecies Growth

(a) and (b)

We follow the steps.

(c) and (d) and (e)

In the 1987-1998 period, average annual growth rates are all negative. from around the 10th percentile, the negative growth rate is similar (roughly -4). However, from 0 to the 10th, the absolute value of average annual growth is smaller, suggesting their consumption did not decrease as much as others. This is specially true for the 2nd and 3rd peecentiles where the graph picks. In the 1998-2008 period, average annual growth rates are mostly negative. Their absolute value increase from 0 to 15th percentile, roughly stagnate from 15th to 95th, and again increase and even become positive at the end. In this case, lower percentiles get poorer, the middle stagnates, and the higher percentiles increase. If we measure inequality in consumption, it is very likely that during the 1998-2008 period, it increases.

5 Lorenz curves (4 points)

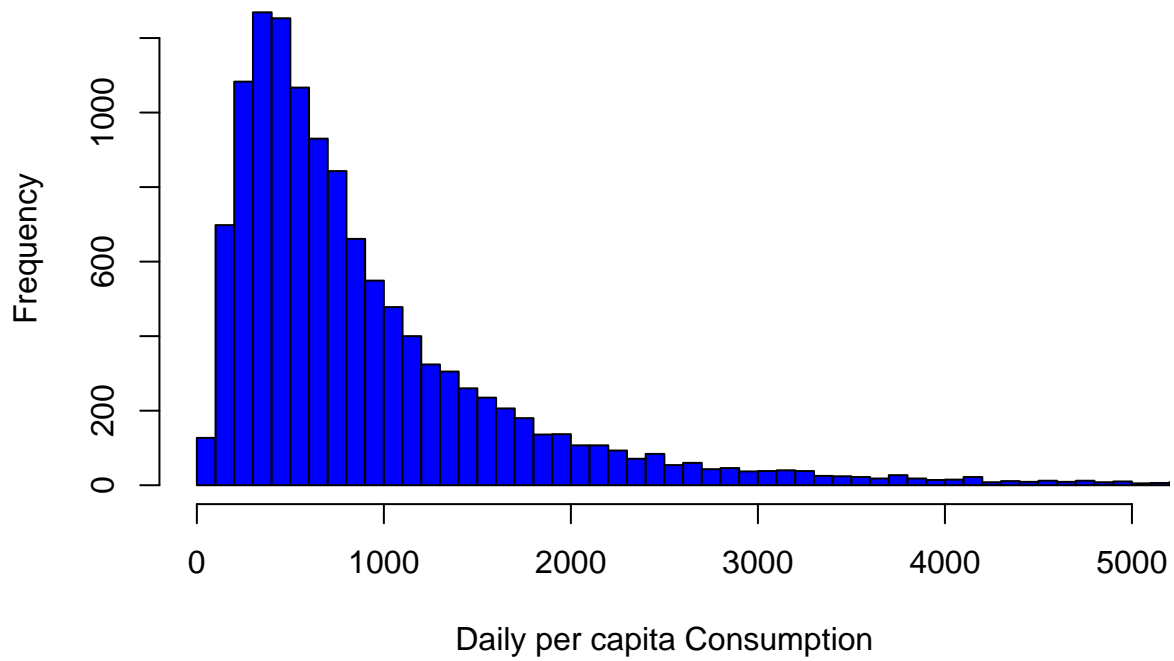
(a) and (b)

Th area bellow the Lorenz curve has a positive relationship with inequality. We use it in the computation of the most famous measurement of inequality, namely the Gini coefficient. In every period, the North has been more unequal than the south. Also, as we saw above, higher percentage of the population are in poverty in comparison to south. Thw lorenz curve of the south is almost the same as the curve at the national level. However, the curve of the north changes during the period. We see a reduction in inequality from 1987 to 1998, but a large increase from 1998 to 2008. This is largely due to the civil war in that period.

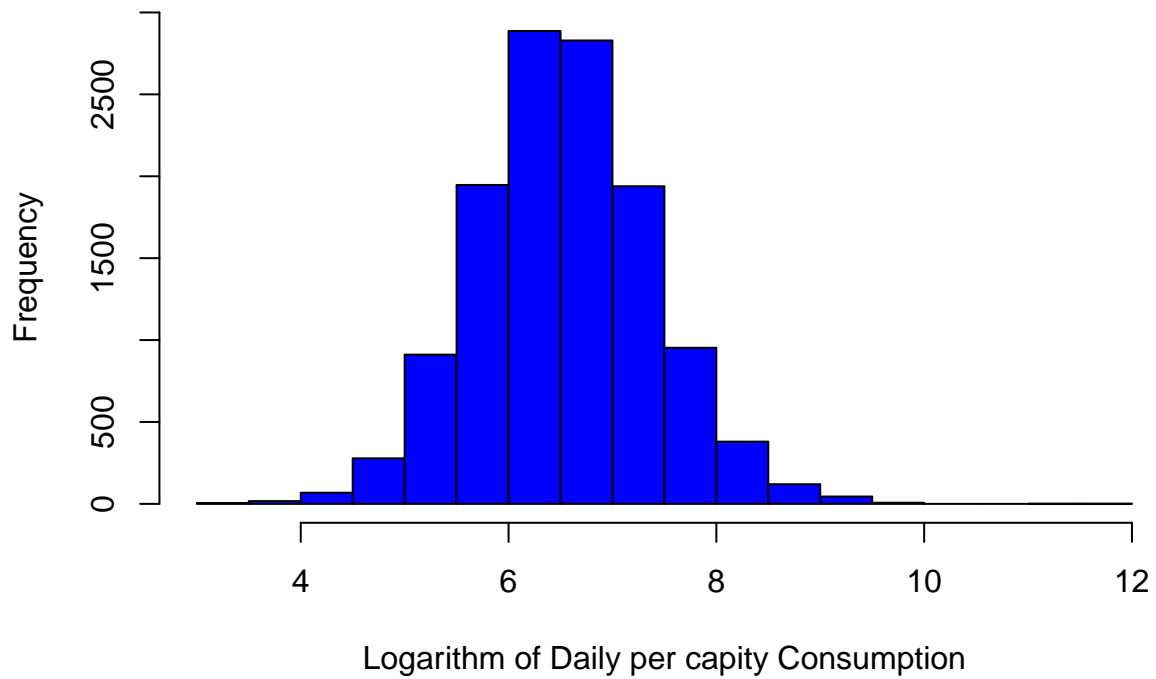
The curves do not cross at the national level. if we look at the last graph, we see that the the lorenz curves are extremly close to each other. Therefore, we can say that the level of inequality in 1987, 1998 and 2008 are the same.

Part 2: Graphs and Tables

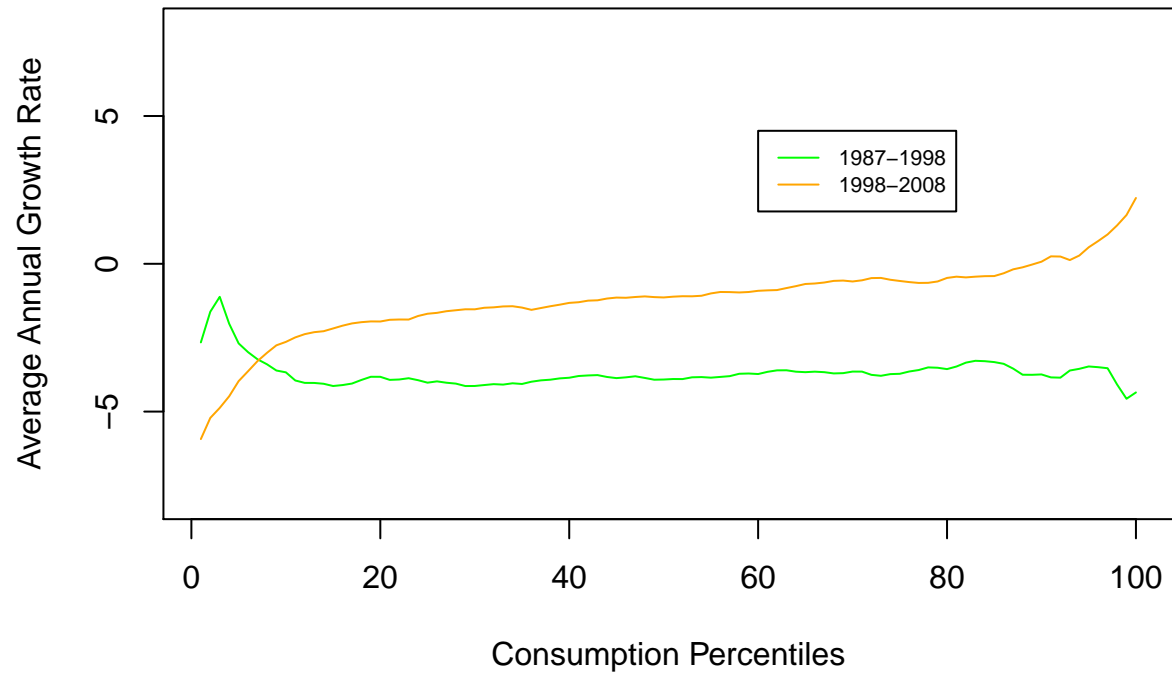
Individual Consumption Distribution 2008



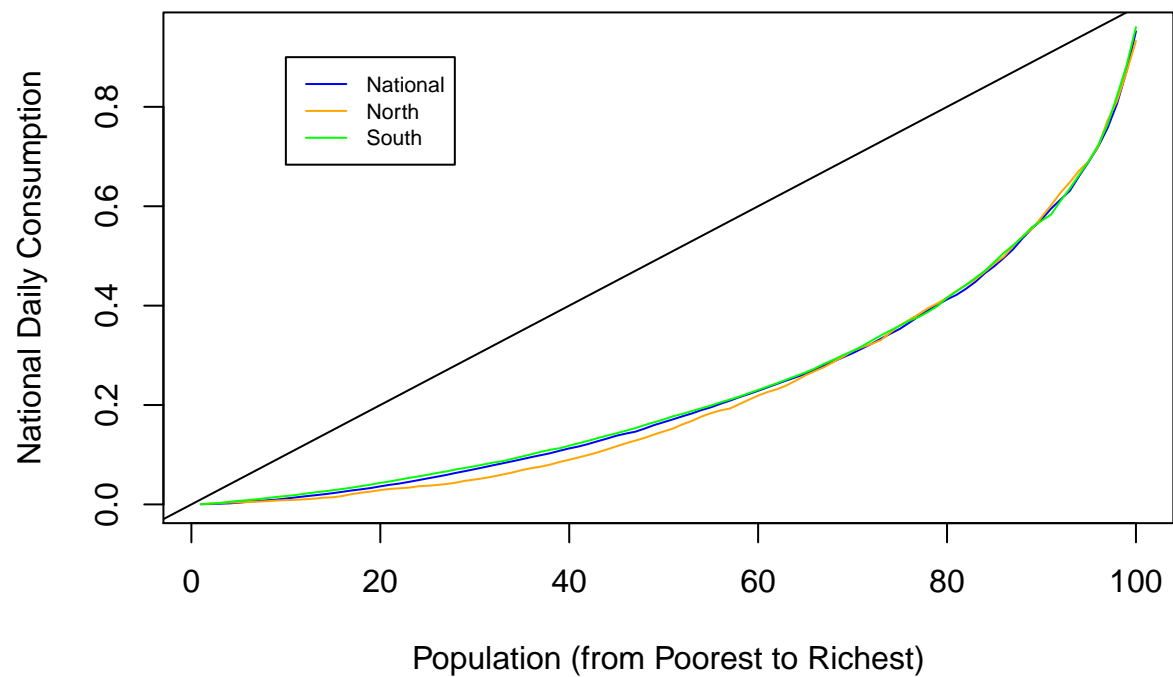
Log Individual Consumption Distribution 2008



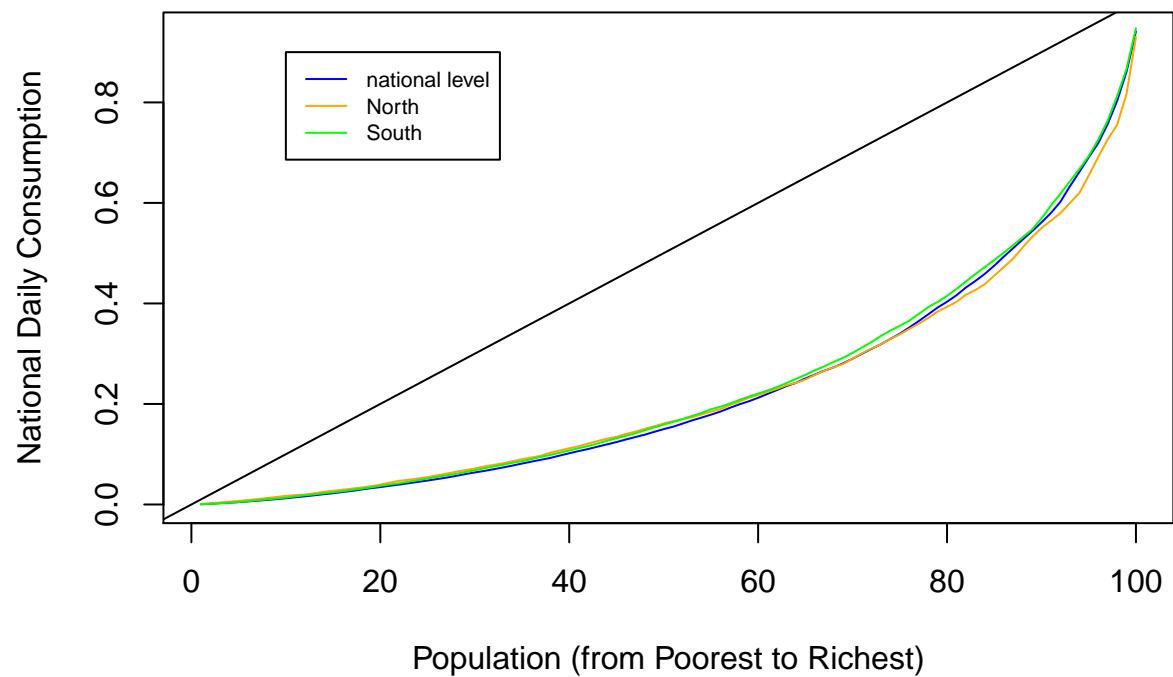
GICs for 1987–1998 and 1998–2008



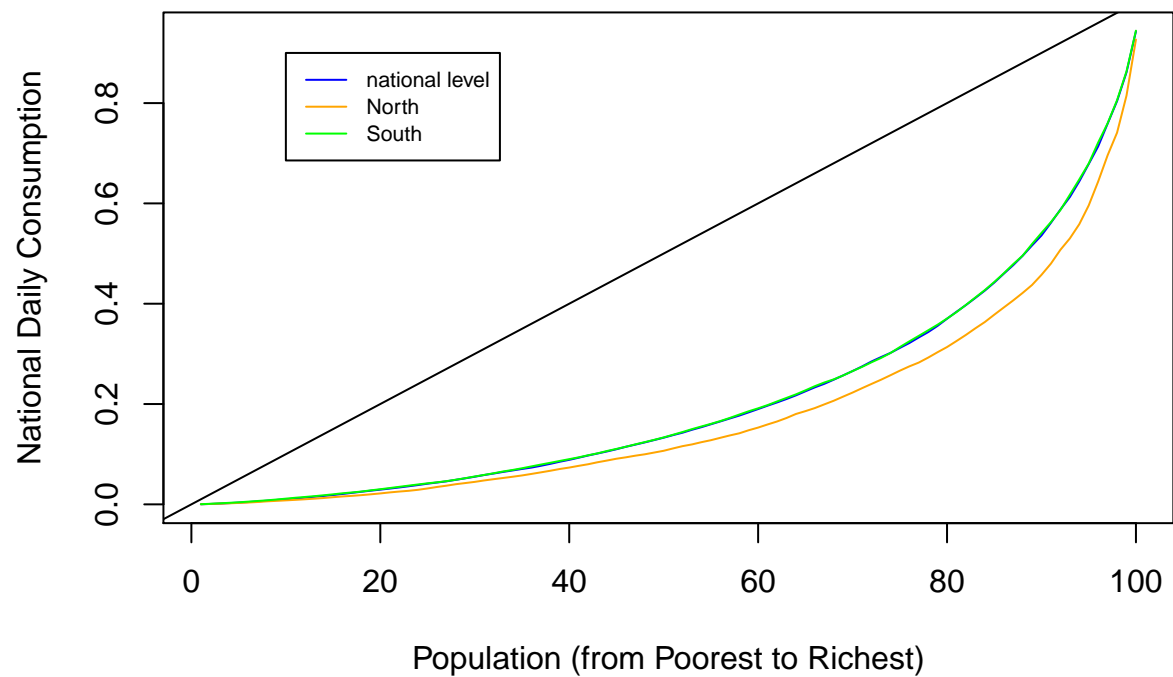
Daily per Capita Consumption–Lorenz Curves 1987



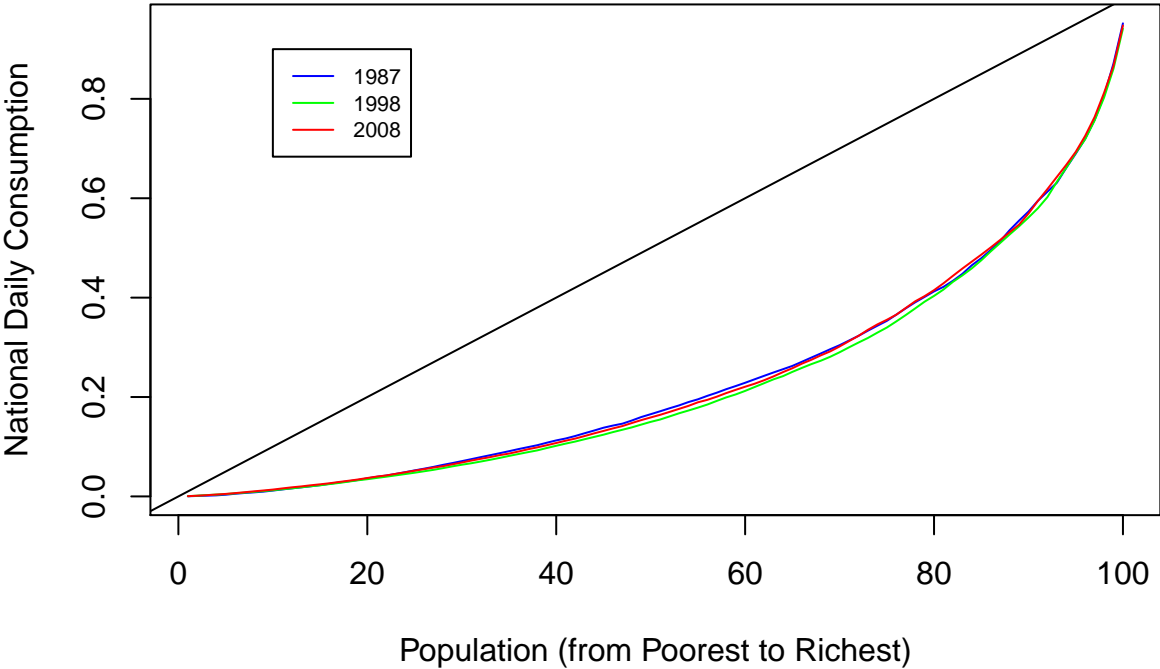
Daily per Capita Consumption–Lorenz Curves 1998



Daily per Capita Consumption–Lorenz Curves 2008



Daily per Capita Consumption–Lorenz Curves in 1987, 1998 and 200



	1987	1998	2008
North			
Food	0.72	0.73	0.61
Non-food	0.28	0.27	0.39
number of observations	396	969	3597
South			
Food	0.63	0.66	0.55
Non-food	0.37	0.34	0.45
number of observation	1188	31182	8791
Total Number of Observations	1584	4115	12388

		1987		1998		2008	
		North	South	North	South	North	South
PPP	L	72.90	100	68.91	100	66.38	100
	P	75.15	100	70.7	100	67.86	100
	F	74.01	100	69.8	100	67.12	100
CPI	L	100	100	180.18	192.51	240.62	255.77
	P	100	100	180.08	192.50	240.76	255.81

		1987			1998			2008		
Poverty Line		North	South	National	North	South	National	North	1998	2008
237	Price levels and preferences differ	38.11%	20.91%	25.11%	67.78%	45.79%	50.92%	75.23%	51.46%	56.76%
	Otherwise	38.11%	29.91%	25.11%	68.55%	45.79%	51.10%	73.82%	51.46%	56.44%
	Number of observations	396	969	3597	396	969	3597	396	969	3597
474	Price levels and preferences differ	76.16%	64.05%	67.01%	93.97%	80.82%	83.88%	94.07%	82.55%	85.11%
	Number of observations	1188	3182	8791	1188	3182	8791	1188	3182	8791
	Total number of observations	1584	4151	12388	1584	4151	12388	1584	4151	12388