Norwegian School of Economics
Bergen, “Autumn 2016”

Income Distribution and Income Inequality in Norway in 1930

Pushkar Singh

Supervisor: Kjell Gunnar Salvanes

Master thesis, MSc., ECN

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

Income inequality is one of the most widely researched topics in economics. Norway and other Scandinavian countries are amongst the most equal countries in world and economists love studying historical inequality in these countries to see whether they were always equal or inequality decreased over the years. Norwegian economists and especially Rolf Aaberge have extensively studied income inequality in Norway from late 19th century to present times using tax records. This thesis studies income distribution and inequality in Norway in 1930 using an extensive unpublished dataset from Statistics Norway archives. The Norwegian welfare model was started in mid 1930s and hence the period analysed in this thesis is before Norway embarked on its path of income equality. Interestingly no one has studied income inequality across Norway in 1930 in detail and hence this thesis fills a gap in current research.

This thesis uses pre tax income from the tax registries for all tax payers from every kommune (municipality) in Norway in 1930 to estimate income distribution and inequality. It shows income distribution for the whole country, for different population subgroups such men and women, rural and urban areas and important rural and urban kommunes using bar charts and kernel density estimation. Income inequality for the entire country and subgroups is calculated using measures such as Gini coefficients and top income shares. Top 1%, 5%, 10% and 50% and bottom 50% income shares among subgroups are compared to find out how income was divided among top and bottom percentiles in every subgroup. Gini coefficients are used to compare overall inequality among subgroups. Inequality in Norway is also compared with other Western countries in the period around 1930.

The results show that Norwegian cities were richer and more equal than rural areas in 1930. Although men were richer than women, there was no big difference in inequality between men and women. The average Gini coefficient of Norway in 1930 was 0.522, much more than Norway's current Gini coefficient. Even though Norway was much more unequal in 1930 than it is today, inequality varied a lot across different parts of the country and population subgroups. Men living in urban areas were the richest and the most equal subgroup. Rural women were the poorest subgroup while rural men were the most unequal subgroup. Urban population was 31% of the entire country and urban areas were richer and more equal than rural areas. Income distribution also varied across cities and villages and some urban kommunes like Oslo and rural kommunes like Aker were much richer than other urban and rural kommunes. Finally, top 1% income share in Norway in 1930 was much less than most other Western countries while income shares of P 90-99 percentile in Norway and other Western countries were similar. As a result, Income share of top 1% contributed less to inequality in Norway as compared to other Western countries.

Preface

This thesis represents the completion of a two year master's programme in Economics at Norwegian School of Economics (NHH), Bergen.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Kjell Gunnar Salvanes at economics department, NHH for all his help and encouragement, which has been far beyond my expectations. I could not have asked for a better supervisor. I am fortunate that both of us are interested in income inequality. I remember him being excited when I first met him and suggested income inequality as my thesis topic. He advised studying Norway in 1930 because he had access to an exclusive unpublished tax database that is still not in the public domain. Not only he provided me with the database, he took a personal interest in my project. Whenever I was stuck, he suggested new approaches. He replied promptly to all my emails at odd hours, gave great feedback and always took out time from his busy schedule to meet me. There is no way I could have completed this thesis without his guidance and help. A big thanks to him!

Second, I would like to thank Dr. Aline Butikofer at economics department, NHH for teaching me labour economics and introducing me to Dr. Salvanes. She is one of the best teachers I had in my life and I think her course in labour economics is the best taught course at NHH. She talked a lot about income inequality and its rise over the 20th century during her course that motivated me to write my master thesis about income inequality.

I would also like to thank Dr. Thomas Piketty, eminent French economist for inspiring me to study income inequality. Last year I read his book "Capital in the twenty-first century" and it has become one of my favourite books. I was always curious about income inequality and in his book, I found answers to most of my questions. After reading his book, I read his research papers and got amazing insights on inequality and how its trends.

Last but not least, I would like to thank my family and friends,both back home in India and all over the world, for their support. When I first told them that I was quitting my banking job and going to far away Norway to study economics, all of them were amused. But they encouraged me to follow my dream and I would not here at NHH without their support. I use this opportunity to say how grateful I am to have such wonderful people in my life.

Any accuracies or errors in this thesis are my responsibility alone.

Pushkar G. Singh

December 2016

Contents

List of Figures

List of Tables

1 Introduction

This master thesis summarises income distribution and inequality in Norway in 1930 using tax records data. Income inequality in Norway and other Western countries are of considerable interest and have been studied by multiple economists in the last 2 decades [(see for example Aaberge, Björklund, et al. 2002); (Aaberge and Atkinson 2010); (Atkinson and Salverda 2005); (Atkinson, Piketty, and Saez 2011)]. The first economist to study income inequality using tax records was Simon Kuznets who tried to find causes in long term change of personal fortune in USA (Kuznets 1955). Kuznets found an empirical relationship between changes in inequality and economic growth and came up with his well-known inverted U hypothesis. Since then economists all over the world have tried to study income inequality using tax records, the most well known of them being Thomas Piketty whose book "Capital in the twenty-first century" has become an important work in understanding how distribution of income and wealth changed in Western countries over the past century. Piketty sketches the evolution of inequality for Western countries since the beginning of industrial revolution and shows that private wealth was much greater than national income in 18th and 19th century and was concentrated in the hands of rich families. (Piketty 2014).

Income inequality in Norway is especially interesting because Norway is one the most equal countries in the world. It had an extremely low Gini coefficient[1] of 0.252 in 2013 (OECD 2013). Among OECD countries, only Iceland has lower levels of income inequality than Norway (OECD 2013). This makes studying historical income inequality in Norway interesting because it tells us about Norway's journey of becoming one of the most equal countries in the world. The question is when did Norway become so equal? Was it always so equal or it eventually became more equal because of the welfare model that the Norwegian state adopted in 1930s? Availability of municipal and central government tax records dating from 1875 make it possible to study changes in income inequality in Norway since late 19th century. Aaberge and Atkinson (2010) explore the long-run changes at the top of income distribution in Norway from 1875 to 2006. They construct an income series using municipal and central government tax records to trace the evolution of the top of the income distribution over a period when Norway industrialised and became income rich (see table 9.1 Aaberge and Atkinson 2010, page 454). Others have studied different periods and compared income inequality in Norway with other countries. For example, Aaberge, Bjorklund and others (2002) compare income inequality and income mobility in Scandinavian countries including Norway with USA from 1980 to 1990 (Aaberge, Björklund, et al. 2002). Aaberge and Atkinson (2016) explore a new series on the distribution of income in Norway as a whole spanning the period from 1875 to 2013. The use sources that provide informa-

[1] The Gini coefficient is an inequality measure and discussed in detail in this report

tion about the incomes of individuals and their families, where income is the total from all sources (earnings, investments and transfers) before deduction of tax. They ignore the data on the distribution of people by occupational groups that attributes them an average income or wage (Aaberge, Atkinson, and Modalsli 2016).

This thesis differs from these analyses of income in Norway because it specifically studies income distribution and inequality in 1930. We use an extensive unpublished data set from Statistics Norway's archive covering the whole population on pretax income from the tax registries from each municipality in Norway. These numbers as averages has only been reported in the census 1930 at the county level, but now we have the data behind these county tables at the kommune[2] level (Norway Census 1930). Hence this thesis fills a gap in research on income distribution and inequality in Norway because no one has analysed income inequality in Norway in 1930 in detail. Income distribution in 1930 is interesting because 1930 was the period between the two world wars and the start of the Great Depression. Norway was largely unaffected during World War I and it was not an industrialised or rich economy by European standards before or after World War I. Although industrialisation in Norway started in 1870, the GDP growth rate was not high until 1930. Norway started experiencing high GDP growth in the beginning of 1930s and this growth continued until 1970. Oil was discovered in Norway in 1969 and this discovery propelled the country to an entirely different income trajectory. So Norway started its transformation from a poor European economy to one of the richest economies in the world in 1930.

To be able to study income inequality in a country and various population subgroups, we first need to to know income distribution in the country. So we start our analysis by measuring income distribution in Norway, across rural and urban areas and for men and women in 1930. Then we compute income inequality using some well-known indicators such as Lorenz curve, Gini-coefficient and income share of top 1%, top 5%, top 10%, top 50% and bottom 50%. We also compare income inequality in Norway with other countries in 1930 using top income shares.

This thesis is structured as follows. Section 2 reviews the economic literature about income distribution and inequality across the world in the period before and after World War II. It also tells us about economic history and income distribution in Norway in late 19th and early 20th century. Section 3 discuss in detail the data used in this paper. It also explains the methodology used to find income distribution and inequality. Results of the analysis are presented in section 4 while section 5 discusses these results and presents key findings. Finally, section 6 summarises the paper and ends with concluding remarks.

[2]Municipalities are called kommunes in Norway.

2 Literature and Background

This section gives a brief discussion about other published research on income distribution and inequality across the world. Income distribution in first half of twentieth century in USA, UK, Netherlands, Canada, France, Germany and Norway is explored. Economic history and income development in Norway between the two world wars and especially in the years before 1930 is also reviewed.

2.1 Income distribution across the world before and after the World War II

The first well known endeavour to study economic growth and income inequality was done by Simon Kuznets in his seminal paper "Economic Growth and Income Inequality". Kuznets (1955) attempts to understand long term changes in personal distribution of fortune. He relates economic growth of a country to income inequality and discusses the causes behind rise or fall in income inequality. He uses the data from pre-tax incomes units for USA and UK in his analysis. He reports that share in income before direct taxes of top 20% in USA in 1929, just before the start of the Great Depression, was 55% while the share of top 5% was 31%. He then proclaims that the share of top 20% declined to 44% in the after war period while the share of top 5% declined to 20%. This decline in top income lead to an increase in share of bottom 40% from 13.5% in 1929 to 18% in years after the World War II. According to his analysis, share of top 5% of tax units in UK declined from 33% in 1929 to 24% in 1947 while the share of bottom 85% increased from 46% to 55% during the same period (Kuznets 1955).

Piketty and Saez (2003) use tax returns data published annually by the Internal Revenue Service (IRS) since the income tax was instituted in 1913 in USA, as well as the large micro-files of tax returns released by the IRS since 1960. They concur with Kuznets (1955) and state that the top income shares fell during the 1914 to 1945 period due to decline of capital income from the Great Depression. Piketty and Saez (2003) think that Kuznets underestimates the top income share, especially for top 1%, in USA because he did not had access to IRS micro-files that led to biases in his estimates due to his methodological assumptions. They estimate income share in USA in 1930 for top 10% at 43% and for top 5% at 31%. This is same as Kuznets (1955) estimates. Similarly, income share of top 10% declines to 34% and top 5% to 24% in 1950 much in line with what Kuznets (1955) says. Interestingly more than half of the decline in the income share of the top 10% in this period is borne by the top 1% who saw their income share decline from 16% to 11%. Authors attribute this decline to destruction of "capital and businesses", which constituted most of the income of top 1%, in the Great Depression and World War II period (Piketty and Saez

2003).

Atkinson and Salverda (2005) use tax data from published tabulations, mostly from the Annual Reports of the Commissioners of Her Majesty's Inland Revenue or in the more recent years from Inland Revenue Statistics to examine evolution of top income in the UK over the 20th century. They manage to calculate shares in total before tax income for 0.1%, 0.05% and 0.01% of the population in 1930 but not for 10%, 5% and 1%. They give estimates for the year 1937 when the income shares of 10%, 5% and 1% stood at 38%, 30% and 17% respectively. Their next estimate is for the year 1949 when income shares of top 10%, 5% and 1% declined to 32%, 23% and 11% respectively. Similar to USA, the decline in income share for the top 10% was mostly accounted by top 1% whose share decreased by 6% in UK as compared to 5% in USA (Atkinson and Salverda 2005).

Atkinson and Salverda (2005) use data from administrative records of tax authorities in the Netherlands to examine evolution of top income shares. They use data from the income tax tabulations, the income distributions based on the income tax data published by Statistics Netherlands (Centraal Bureau voor de Statistiek: CBS) and the Income Panel Survey (Inkomenspanelonderzoek: IPO), a source of micro-data that is also maintained by CBS for the period starting in 1977. They find out that the income shares of top 10%, 5% and 1% in 1930 was 43%, 32% and 17% respectively. These income shares declined to 36%, 27% and 12% respectively in 1950. Like USA and UK, decrease in income share of the top 1% was responsible for most of the decline in income share of top 10% in Netherlands between 1930 and 1950 (Atkinson and Salverda 2005).

Saez and Veall (2005) estimate top income shares in Canada between 1920 to 2000 from personal income tax return statistics compiled annually by the Canadian federal taxation authorities. They do not give top 10% income share for any year before 1941 but they estimate that the income shares of top 5% and top 1% was 33% and 16% respectively in 1930. They estimate top 10% income share at a relatively high 38% in 1950 and the income share for top 5% and top 1% at 25% and 11% respectively in 1950. (Saez and Veall 2005).

Thomas Piketty use three major types of data sources namely data from income tax returns (1915–98), data from wage tax returns (1919–98), and data from the inheritance tax returns (1902–94) to construct an income series for France between the years 1901 to 1998 in his book "Income, Wage, and Wealth Inequality in France, 1901–98". He estimates shares in total before tax income for 10%, 5% and 1% in 1930 to be around 41%, 30% and 15% respectively. This share declines to 32%, 22% and 9% respectively in 1950. Like other countries, decline in income share of top 1% in France formed a major part of decline in income share of top 10% (see Piketty 2007, pages 43-81).

Dell (2007) analyses income inequality in Germany using tax returns statistics compiled by the successive German fiscal administrations over the twentieth century. He finds that share of top 1% in Germany showed a very minute increase from 11.2% in 1928 to 11.6% in 1950. On the other hand, top 5% and 10% income shares increased from 23% and 32% respectively in 1929 to 25% and 34% respectively in 1950. The change in income shares of top 1%, 5% and 10% in Germany followed a very different course between 1930 and 1950 as compared to USA, UK, Netherlands, France and Canada (Dell 2007).

From the above research, it is evident that income share of top 10%, top 5% and top 1% declined in all these countries (with the exception of Germany where it remained constant or increased) between 1928-1937 and 1950. What is interesting is how similar this decline was for all income groups. For example, in this period, share in pre-tax income of top 5% in USA and UK as well as in Netherlands and Canada decreased from around one-third to one –fourth of total income.

Aaberge and Atkinson (2010) use Norwegian municipal and central government income tax records to construct an income series dating from 1875 till 2006. They use tabulations of the distribution of income as assessed for tax purposes, giving the number of income recipients and total amount of income by ranges of assessed income from 1875 till 1966. From 1967 to 2006, they use tax micro-data from tax files available at Statistics Norway. In 1929[3], income share of 10%, 5% and 1% was 41%, 28% and 13% respectively. These income shares dropped down to 34%, 22% and 9% respectively in 1950. As we can see that, although top 1% in Norway also saw their incomes decline like other countries, the absolute share of top 1% in Norway was much less than USA, UK, Netherlands and Canada both in pre-World War II period and 1950 (Aaberge and Atkinson 2010). We show top income share of 1%, 5% and 10% in all these countries in table 2.1.

[3]There is no series for 1930

Table 2.1: Top income shares in select Western countries

	10 %	5 %	1 %
USA (1930)	43.07 %	31.18 %	16.42 %
UK (1937)	38.37 %	29.75 %	16.98 %
Netherlands (1930)	43.02 %	32.41 %	17.15 %
Canada (1930)	NA	32.74 %	16.10 %
France (1930)	41.08 %	30.14 %	15.31 %
Germany (1928)	32.20 %	22.60 %	11.20 %
Norway (1929)	41.32 %	28.25 %	12.57 %
USA (1950)	33.87 %	23.87 %	11.36 %
UK (1949)	32.25 %	23.39 %	11.47 %
Netherlands (1950)	36.74 %	26.16 %	12.05 %
Canada (1950)	38.24 %	25.45 %	10.88 %
France (1950)	31.97 %	21.62 %	8.98 %
Germany (1950)	34.40 %	24.90 %	11.60 %
Norway (1950)	34.10 %	22.09 %	8.76 %

Source: (Piketty and Saez 2003), (Atkinson and Salverda 2005), (Saez and Veall 2005), (Piketty 2007), (Dell 2007), (Aaberge and Atkinson 2010)

Another widely used indicator to measure income inequality is the Gini coefficient[4]. The Gini coefficient, also known as a Gini ratio or a normalised Gini index, measures the inequality among values of a frequency distribution (for example, levels of income). A Gini coefficient of zero expresses perfect equality where everyone has the same income, whereas a Gini coefficient of 1 (or 100%) expresses maximal inequality where one person has all the income.

Bourguignon and Morrisson (2002) analyse inequality amongst world citizens during 1820 – 1992. Unlike most researchers, they do not use the tax records to estimate the income distribution for all the major countries but 3 different types of data: real GDP per capita expressed in PPP dollars, population and distribution of income summarised by 9 decile shares and 2 vintile shares. They update previous works on world inequality and put a historical perspective in their analysis. Interestingly 1929 is one of the years for which they analyse inequality across the world. They estimate the Gini coefficient of the world in 1929 to be around 0.616 and declare that it kept increasing over the 20th century (see Bourguignon and Morrisson 2002, pages 729 and 732).

Paukert (1973) mentions that unique historical records for Norway make it possible to cal-

[4]See section 3.3.1 for the methodology behind calculating the Gini coefficient.

culate the Gini coefficients for eight Norwegian cities for ten year intervals between 1840 and 1960. He uses the data from Lee Soltow's book "Toward income inequality in Norway" to calculate the Gini coefficients for men in 8 Norwegian cities.

Table 2.2: The Gini coefficients for men in 8 cities, Norway 1930

City	County	Gini
Sarpsborg	Østfold	0.365
Fredrikstad	Østfold	0.421
Halden	Østfold	0.434
Moss	Østfold	0.420
Kristiansand	Vest-Agder	0.406
Mandal	Vest-Agder	0.432
Flekkefjord	Vest-Agder	0.398
Farsund	Vest-Agder	0.389

Source: (Paukert 1973) (Soltow 1965, page 17)

Paukert (1973) claims that although this data indicates income distribution only in limited areas of Norway, it is much more accurate than comparable income data available for other countries such as UK in that period. He also measures the Gini Coefficients for USA using distribution of family personal income before tax between 1929 and 1958. He does not give the Gini coefficient for USA in 1930 but mentions that it was 0.47 in 1935-36. The Gini coefficient of other countries in the period before World War II is not so well documented but the Gini coefficients of USA and Norway give us a fair idea about the income inequality in the world in the 1930s. The Gini coefficient of the world in 1929 was 0.616 (see Bourguignon and Morrisson 2002, page 732), much more than the Gini coefficient of Norway and USA in the same period. We have seen that the income shares of top 10%, 5% and 1% in USA, UK, Netherlands, Canada, France, Germany and Norway in 1930 were close to each other. So it is reasonable to assume that Gini coefficients of these countries were also in the similar range even though we do not have estimates. These Gini coefficients are much less than the Gini coefficient of the entire world in years around 1930 indicating much higher levels of income inequality in rest of the world as compared to Western countries.

2.2 Economic history and income development of Norway in the late 19th century and early 20th century

Norway was a pre-industrial society with agriculture and fishing as its main industries in the first half of 19th century. In the late 1870s just one seventh of Norwegian population lived in urban areas. Although Norway was a non industrial economy in 1870, it had a vibrant shipping sector. Francis Sejersted estimates the share of shipping, fish and timber to be around 12% of GDP in 1870 (Sejersted 1992). Aaberge and Atkinson (2010) mention predominance of shipping sector in Norway as one of the reasons behind the high income share of 0.5% in 1875. Norway started industrialising in the second half of 19th century and developed into an industrial economy by 1950. There were a few recessions and economic downturns in between such as few years of recession around 1880, Kristiania crash in 1899, a mild recession in 1908 and then years in the aftermath of the Great Depression. The Great Depression was milder and shorter in Norway than most other Western countries, i.e. GDP growth and prices fell less and recovery was faster. The economic growth during 1870 till the Great Depression in 1930 was accompanied by a shift in population from rural to urban areas. As a result around 31% of adult population[5] was living in the urban areas in 1930 (Norway Census 1930). This shows that Norway became much more urbanised in the period between 1870 and 1930. Even though Norway experienced economic growth, it remained one of the poorest countries in Europe and this led to massive emigration to USA in the late 19th century and early 20th century. Emigration among youth was very high in the late 19th century, it slowed down in the early 20th century and came almost to a standstill by the time World War II started. This led to demographic change and the proportion of people at least 20 years or more in the population increased from 55% in the period between 1875 - 1892 to 62% in the period between 1923 - 1939 (see Aaberge, Atkinson, and Modalsli 2016, page 28).

The composition of Norwegian economy changed much in the period between 1875 and 1939 as the share of secondary and service sectors increased while that of primary sector declined. Primary sector's contribution to the GDP decreased from 33% to 17% in the period between 1875 and 1939 while the share of secondary sector increased from 25% to 33% and the share of service sector increased from 42% to 50% (see Aaberge, Atkinson, and Modalsli 2016, page 28). Aaberge, Atkinson and Modalsli (2016) try to find relationship between evolution of income inequality and changes in rate of emigration or economic growth between 1875 and 2013. They find no clear relationship between change in income inequality and changes in rate of emigration or GDP growth. For example, inequality rose between 1875 and 1892 and declined between 1892 and 1914 while emigration was high in

[5] 15-year-old and above

this entire period. Similarly, although Norway experienced high economic growth between 1892 and 1939, inequality decreased in 1892 - 1914 while it increased during 1923 - 1939 period. This makes 1930 an interesting year to study income inequality in Norway because inequality was on rise during this period after declining in late 19th and early 20th century.

3 Methodological Approach

This section explains the methodological approach used in this thesis. First, the data used in this thesis is described. Then income inequality and various measures of income inequality are touched upon. Finally, the calculation of various income inequality measures used in our analysis is explained in detail.

3.1 The income tax data in Norway

We use an extensive unpublished data set (Norway Census 1930) covering the whole population on pretax income from the tax registries from each municipality in Norway from Statistics Norway's archive. These numbers as averages has only been reported in the census 1930 at the county level, but now we have the data behind these county tables at the kommune level. This data set has pre-tax income of around 900,000 tax-payers in 742 kommunes (municipalities) across 20 fylker (counties) that existed in Norway in 1930[6]. Bergen and Oslo are 2 counties that are not subdivided into different kommunes. Hence, Bergen and Oslo are both counties and kommunes. Three kommunes: Aker (Akershus county), Bergen and Oslo are further subdivided into various neighbourhoods. Each kommune is classified as either rural or urban with 675 kommunes being rural and remaining 67 kommunes as urban. Hence the number of tax payers in rural kommunes is much more than that in urban kommunes.

Each kommune has data for number of people at least 15-year-old, number of people at least 15-year-old who are not in tax records and number of tax payers at least 15-year-old. Each kommune has separate observations for men and women. If we aggregate the data for all the kommunes, we get the total number of people at least 15-year-old and the total number of tax payers at least 15-year-old in Norway in 1930. The total adult population of Norway in 1930 was 1,996,490, out of which 887,107 were registered in the tax records. Our analysis is based on the income record of these 887,107 tax payers because we don't have income information about the people not in tax records. Another interesting feature of the data is that the number of female tax payers is much less than that of male tax payers. One reason could be that in 1930 married women were considered in the same tax unit as their husbands. So it is fair to assume that all female tax payers in the data were unmarried women. This partly explains why registered tax payers are less than half of adult population because married women are not counted as individual tax payers. Among the registered tax payers, around 31% are from urban kommunes while the rest 69% are from rural kommunes. This indicates that Norway was primarily a rural country as late as 1930. Table 3.1 shows the number of

[6]City of Bergen was county number 13 in 1930 and got later merged into Hordaland, county number 12, in 1972.

registered tax payers in the entire country and number of male & female and rural & urban tax payers.

Table 3.1: Overview of tax payers, Norway 1930

	Total	Rural	Urban
Men	660,674	473,572	186,734
Women	226,433	135,062	91,739
Both sexes combined	887,107	608,634	278,473

Source: (Norway Census 1930)

The data set covers tax payers incomes across 20 Norwegian counties in 1930. Table 3.2 shows the names and numbers of all 20 counties in Norway in 1930 and number of people at least 15-year-old and number of tax payers at least 15-year-old in each county.

Table 3.2: Adult population and tax payers in all counties, Norway 1930

County Number	County Name	Number of people at least 15-year-old	Number of tax payers at least 15-year-old
1	Østfold	120,232	56,004
2	Akershus	155,115	65,025
3	Oslo	210,295	95,586
4	Hedmark	106,372	50,365
5	Oppland	95,499	44,217
6	Buskerud	102,239	43,010
7	Vestfold	97,264	44,057
8	Telemark	91,246	37,765
9	Aust-Agder	53,336	22,197
10	Vest-Agder	58,288	24,532
11	Rogaland	119,832	48,460
12	Hordaland	112,771	50,456
13	Bergen	74,858	41,895
14	Sogn og Fjordane	65,449	33,492
15	Møre og Romsdal	115,958	55,203
16	Sør-Trøndelag	125,024	55,422
17	Nord-Trøndelag	67,041	33,362
18	Nordland	126,558	47,748
19	Troms	64,801	23,917
20	Finnmark	34,312	14,394
	Total	1,996,490	887,107

Source: (Norway Census 1930)

We can see that Oslo county, which was also a kommune, had the biggest population as well

as highest number of tax payers among all counties in 1930. Bergen, although not a big county, was the second biggest kommune after Oslo. Other counties with population of over 100,000 adults were Østfold, Akershus, Hedmark, Buskerud, Rogaland, Hordland, Møre og Romsdal, Sør-Trøndelag and Nordland. We analyse income distribution of some important kommunes from these big counties in detail in this thesis.

The data set does not record the individual annual income of every tax payer in Norway. Instead, all tax payers are grouped into one of several income groups as per his or her annual income. So we have grouped data for tax payers where each group represents an income range[7] and a value (observation) that tells us the number of tax payers in that income group. The data set has observations for 742 kommunes with every kommune having two observations, one for men and one for women, in each income group. Income observations use NOK (Norwegian Kroner) as unit. Number of income groups vary across kommunes. Some kommunes have more income groups and while other have less. For example, top income group for all rural kommunes is an annual pre-tax income of 20,000 NOK & above. On the other hand, this 20,000 NOK & above income group is divided into 20,000 – 49,900 NOK and 50,000 NOK & above for urban kommunes. Some urban kommunes like Bergen and Oslo have more subdivisions for lower income groups while others have less. To make the data uniform and comparable, we merge some existing income groups to generate new income groups that are common for all kommunes. The end result is that all rural kommunes (total of 675) are divided into 10 income groups while all urban kommunes (total of 67) into 11 income groups. The reason being urban kommunes have 20,000 – 49,900 and 50,000 & above as 2 distinct income groups while rural kommunes have 20,000 & above as one income group.

There are 2 observations in each income group for all kommunes: number of male tax payers and number of female tax payers. When we add both these observations, we get income distribution for all the tax payers at the kommune level. Adding all the kommunes under a county gives us income distribution for the county. Similarly, we can calculate the income distribution for the whole country, for urban and rural areas and for men and women by aggregating kommune observations. We face one minor issue while merging both rural and urban kommunes to generate income distribution for the entire country. All rural tax payers with an annual income of 20,000 NOK & above are represented in the income group 20,000 NOK & above while urban tax payers having the same range of income are divided into 2 income groups. Therefore, when we design the income groups for the entire country, we merge the top 2 income groups (20,000 – 49,900 NOK and 50,000 NOK & above) of the urban tax payers into one income group (20,000 NOK & above). This creates

[7] referred as income group in this report

identical income groups for both urban and rural tax payers and makes it easier to merge rural and urban income distributions to create income distribution for the entire country with 10 income groups. Similarly, income distribution of all male and female tax payers have 10 income groups because rural men and women are divided into 10 income groups while urban men and women into 11 income groups.

3.2 Measuring income distribution

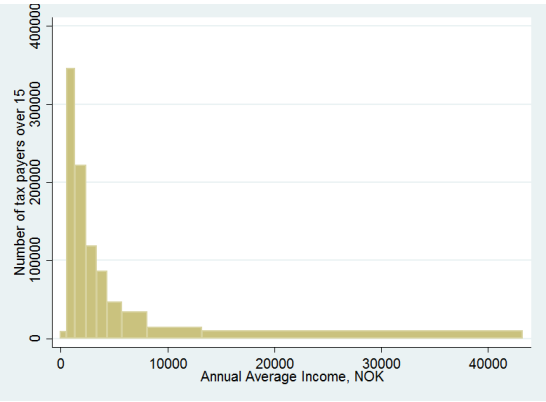
This thesis attempts to show the income distribution in Norway in 1930 and then measure income inequality in Norway in 1930. By income we mean pre-tax income of all tax payers at least 15-year-old in 1930. It is possible that these tax payers had other sources of income that is not reflected in the income tax records. Income after taxes and transfers may also substantially differ from the original pre-tax incomes since high income individuals pay a higher percentage of their income as tax while low income individuals get transfers from the government. Since we don't have any information about these income sources, we use pre-tax income recorded in the census for measuring income distribution.

It is difficult to make any assumptions regarding the distribution of the variable (income in our case) because we don't know whether the distribution is normally distributed or follows any other distribution. Thus, we use non parametric models to analyse the distribution of income. The most common non parametric models to estimate a probability distribution are histogram and kernel density estimation.

3.2.1 Histogram

A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable. Continuous variable could be anything like annual incomes or wealth on people in a country or region. To construct a histogram, first we need to "bin" the range of values (divide the entire range of values into a series of intervals). We use Stata for our calculations that has an option to specify the bin while drawing the histogram. If no bin option is given, Stata calculates bin using $k = \min\{(\sqrt[2]{N}), \frac{10 \ln N}{\ln N}\}$, where N is the (weighted) number of observations. In our case, since we already have the income ranges, we use them as bins. To construct a histogram, we use different income groups as separate bins and then count how many values fall in each bin. The income is already divided in continuous and non-overlapping intervals in our data. However, there is one challenge. There is no upper limit for the top income group. Top income group for rural kommunes is 20,000 NOK & above while for urban kommunes is 50,000 NOK & above. One way to solve this is to assign a value for the upper bracket of the top income group, say 100,000 NOK, and then draw a histogram. If the income intervals are of equal size, a rectangle is erected over the income interval showing the number of tax payers in that income interval. The width of each rectangle will be the same in this case because income intervals are identical in size. Since in our data the income intervals are of unequal size, the width of rectangle represents the relative size of the income interval. Figure 3.1 shows a histogram with income intervals of unequal sizes.

Figure 3.1: An example of a histogram



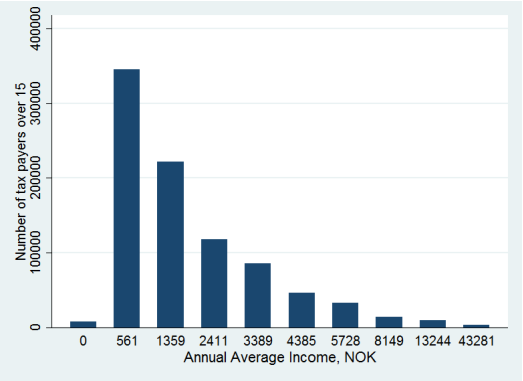
In the above histogram, the width of each rectangle represents the income interval and height of each rectangle show the number of tax payers in that income interval. Although this histogram is informative, we can use another method to show income distribution in a better way. This histogram is missing one important information that we have in our data. It does not show the mean income under each income group. We have total income and hence mean income for all the tax payers under each income group. We believe it is more informative to use mean income because of 2 reasons. First is income groups give us a broad range of incomes rather than specify the actual incomes. Average or mean income on the other hand tells us more about the tax payers in that income group. For example, income group of 10,000 to 20,000 does not tell us whether there are more tax payers with an income of 11,000 or 17,000. On the other hand, if mean income is 13,500, there is a higher chance that more tax payers have an income closer to 10,000 than 20,000. An important property of the mean is that it includes every value in the data as part of the calculation. Mean is also the only measure of central tendency where the sum of the deviations of each value from the mean is always zero. But the average or mean has one main disadvantage: it is predominantly vulnerable to the influence of outliers. Few tax payers with very high income or very low income can influence the mean. Mode is a better option if we want to show the income of most of tax payers but unfortunately we do not have individual incomes or any way to estimate mode. Second reason behind using average or mean income instead of income bracket is that we do not have an upper bracket for the top income group. Using average or mean income saves us from making this assumption.

3.2.2 Bar charts

Now the question is how to represent income distribution using mean incomes. We can not use histogram in this case because now the incomes are non-continuous and discrete. We use bar charts for this purpose because they can be used to represent non-continuous variable. A bar chart also known as bar graph or line graph is a chart that presents grouped

data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. Bar charts use discrete data and one axis shows the specific categories being compared while the other axis shows discrete values. We use vertical bar charts in our report with x-axis showing the mean income in each income group while y-axis representing the number of tax payers in that income group. Figure 3.2 shows a bar chart using mean (average) income instead of income interval on the x axis.

Figure 3.2: An example of a bar chart



The above bar chart is more illustrative than a histogram because it tells us that the tax payers in first income bracket (0 – 900 NOK) have an average income of 561 NOK. So for our data, bar chart gives some additional information over histogram.

3.2.3 Kernel Density Estimation (KDE)

Another popular non parametric model to estimate the distribution is kernel density estimation (KDE). Kernel density estimators approximate the density f(x) from observations on x. Histograms do this, too, and the histogram itself is a kind of kernel density estimation. In KDE, the data is divided into non overlapping intervals, and counts are made of the number of data points within each interval. Histograms as mentioned above are bar graphs that depict these frequency counts. Kernel density estimation has the advantages of being smooth and of being independent of the choice of origin (corresponding to the location of the bins in a histogram).

Definition of Kernel:

Let $x_1, x_2, x_3, x_4, \cdots, x_n$ be an independent and identically distributed sample drawn from some distribution with an unknown density f. We are interested in estimating the shape of this function f. A kernel density estimation is formed by summing the weighted values calculated with the kernel function K. Stata uses the below formula to calculate f_K

$f_K = \frac{1}{qh} \sum_{i=1}^{n} w_i K\left(\frac{x - X_i}{h}\right)$

Here K is a non-negative function that integrates to one and has zero mean and $q = \sum_i w_i$ when the weights (w_i) denote frequencies of each x. If the frequency of each $x = 1$ then $w_i = 1$ for $i = 1, 2, 3, \cdots, n$. K density includes eight different kernel functions namely Biweight, Kosine, Epanechnikov, Epan2, Gaussian, Parzen, Rectangular and Triangular. The Epanechnikov kernel is the default function in Stata if no other kernel is specified and is the most efficient in minimizing the mean integrated squared error.

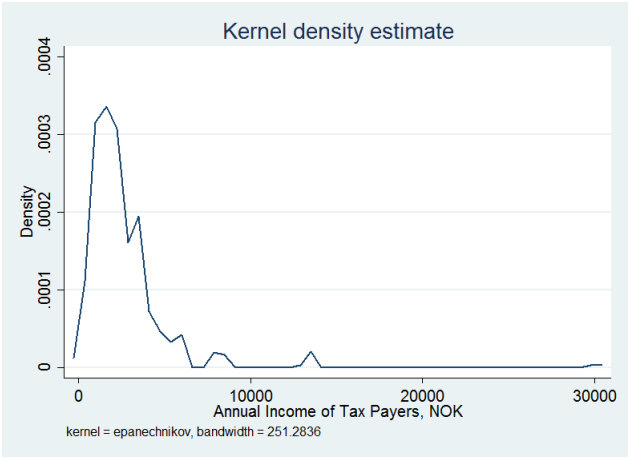
The choice of h (a positive parameter called bandwidth) determines how many values are included in estimating the density at each point. If h is not given, Stata estimates it using

$$h = \frac{0.9m}{n^{\frac{1}{5}}}$$

where $m = \min\left(\sqrt{variance_x}, \sqrt{\frac{interquartile\ range_x}{1.349}}\right)$ where x is the variable for which we wish to estimate the kernel and n is the number of observations.

We use KDE to show the income distribution in some rural and urban kommunes. Figure 3.3 shows a KDE with income range on x axis and density on y axis.

Figure 3.3: An example of a kernel density estimation (KDE)



The above KDE chart tells that most of the distribution lies between 0 and 10,000 NOK and then there are some spikes. The reason has behind these spikes is because of the nature of our data. Our data assigns one common mean income to all tax payers in one particular income category. Since we do not have incomes of individual tax payers, all tax payers in one income group are plotted against the mean income of that income group. First spike shows the tax payers with mean income between 10,000 and 20,000 NOK while the second shows tax payers with mean income around 30,000 NOK.

3.3 Measuring income inequality

Once we know the distribution of income in Norway in 1930, we attempt to measure the income inequality. One important concern in calculating income inequality is the definition of income inequality. We use the same income and population criteria for computing income inequality that we use for estimating income distribution. One of the key challenges in measuring income inequality from income distribution is finding an appropriate metric.

There are several metrics to estimate income or wealth inequality. Suppose $I(x)$ with $x = x_1, x_2, x_3, \cdots x_n$ represents an income inequality measure or index where x_i is the economic value (say income) associated with x agents (tax payers, people etc.). Economic literature tells us that this $I(x)$ should at least have the following 4 properties to be an unbiased and correct estimator.

Anonymity or symmetry: This assumption states that an inequality metric does not depend on the order in which the agent's income is measured. This means that any permutation of agents (tax payers in our case) can be used while measuring the inequality and the result remains unchanged.

$I(x) = I\{P(x)\}$ where $I\{P(x)\}$ is a permutation of $I(x)$.

Scale independence or homogeneity: This property says that the income inequality metric should be independent of the aggregate level of income. If every agent's income in an economy is multiplied by any positive number, then the overall metric of inequality should not change.

$I(\alpha x) = I(x)$ where $\alpha > 0$

Population independence: This property says that the income inequality metric should not depend on the number of agents or the size of the population.

$I(x \cup x) = (x)$ where $x \cup x$ is an union of x with itself

Transfer principle or The Pigou–Dalton principle: This principle says that if some income is transferred from a rich person to a poor person while preserving the order of income ranks, the measured inequality should not increase. In its strong form, this principle says that the measured inequality should decrease.

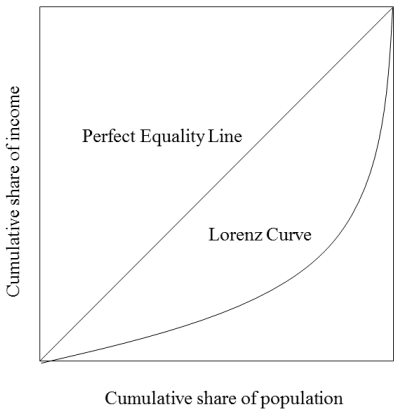
There are some non-mandatory properties of an inequality metric such as non-negativity (greater than or equal to 0) and egalitarian zero (if everyone has same income then the inequality measure is 0). One more desirable property that all commonly used inequality

measures don't satisfy is "subgroup decomposability". Subgroup decomposability states that if a particular economy is decomposed into sub-regions, and an inequality metric is computed for each sub region separately, then the inequality metric for the whole economy should be a weighted average of the regional inequality metric plus a term proportional to the regional inequality metrics[8].

3.3.1 Gini coefficient

The Gini coefficient (Gini ratio or Gini index) is a metric used to represent the income distribution of a nation's residents. The Gini coefficient is widely used to measure inequality in the distribution of income, wealth, expenditures, etc. It satisfies all the 4 mandatory properties of an income inequality estimator. It is an inequality measure that is mostly associated with the descriptive approach to inequality measurement. Gini coefficient gives summary information on income distribution and does not give information on characteristics of income distribution such as location and slope (Bellù and Liberati 2006). The Gini coefficient was developed by Corrado Gini in 1912, building on the work of an American economist Max Lorenz who published a hypothetical way to depict perfect equality - a straight diagonal line on a graph - in 1905. The Gini coefficient represents the income inequality through the Lorenz Curve, developed by the same Max Lorenz. The Gini coefficient plots the proportion of the total income of the population that is cumulatively earned by the bottom x% of the population. The line at 45 degrees, also known as Perfect equality line (straight diagonal line first used by Max Lorenz), represents perfect equality of incomes. The Lorenz curve is a graph used to represent income or wealth distribution by showing the proportion of overall income or wealth assumed by bottom x% of the people. Figure 3.4 shows the 45-degree line (Perfect equality line) and the Lorenz curve against the overall income of population.

Figure 3.4: An example of a Lorenz curve



[8]The Gini coefficient, one of the most widely used inequality metric, does not satisfy this property.

Let X be an income variable with cumulative distribution function $F(.)$[9], density $f(.)$ and mean μ. Let $[0,\infty]$ be the domain of F where $F^{-1}(0) = 0$. Aaberge (2007) defines Lorenz curve $L(u)$ as

$L(u) = \frac{1}{\mu}\int_0^{\mu} F^{-1}$ Where F^{-1} is the left inverse of F. An advantage of the Lorenz curve is that it is concerned with shares of income rather than relative levels of income (see Aaberge 2007, page). One important way the Lorenz curve differs from the decile-specific presentation of income inequality is decile-specific metrics always display mean incomes as fractions of the overall mean income. Since the Lorenz curve represents income distribution by showing the proportion of overall income or wealth assumed by bottom x% of the people and 0% of the population will always have 0% of the income while 100% of the population will always have 100% of the income, $L(0) = 0$ and $L(1) = 1$.

The Gini coefficient is the ratio of the area that lies between the Perfect equality line and the Lorenz curve over the total area under the Perfect equality line. This is same as one minus two times the area under the Lorenz curve. Let us assume the total area under the Perfect equality line as "a", the area between Perfect equality line and the Lorenz curve as "b" and the area under the Lorenz curve as "c". So, we have $a = b + c$ and Gini coefficient G is $G = \frac{b}{a}$.

If we substitute $a - c$ for b, we get $G = (a - c)/a = (1 - c/a)$. Since the area under the Perfect equality line is an isosceles triangle with one side equal to 1, $c = (1/2)$. Thus Gini coefficient is $G = 1 - 2c$, same as one minus two times the area under the Lorenz curve. This can be written as $G = 1 - \int_0^1 L(u)du$.

Gini coefficient ranges from 0 to 1 where 0 is complete equality and 1 is complete inequality. Gini coefficient is also represented as a percentage between 0 and 100. In case of negative wealth or income, Gini coefficient can be theoretically more than 1. The total wealth or income is always assumed to be positive, unless an entire country is in debt, and hence Gini coefficient cannot be negative.

The biggest advantage of using Gini Coefficient to measure income inequality is that it facilitates direct comparison between two populations, regardless of their sizes. In other words, we can directly compare the inequality in male tax payers to the inequality in female tax payers in 1930 even though population varied a lot between the 2 subgroups. The biggest drawback of the Gini coefficient is that although 2 subgroups may have same Gini Coefficient, they can have different inequalities because the two Lorenz curves can have same area

[9] $F(.)$ can either be a discrete or continuous distribution function. Although it is often observed as discrete, we can use a continuous $F(.)$ to make derivations easier when we deal with large samples.

yet different shapes.

3.3.2 Gini coefficient from grouped data

We face one challenge to calculate the Gini coefficient from our data. We do not have individual income observations of all tax payers. Tax payers are grouped under different income groups and hence we need a technique to calculate the Gini coefficient from grouped data. One way to do is to consider average income in each group as one observation and the number of tax payers in that income group as weights. Milanovic (1994) proposes an alternative and simple derivation of the Gini coefficient and then derives a number of coefficients (concentration coefficients) from this Gini coefficient. Aboundoori and McCloughan (2003) modify this formula (Milanovic 1994) and come up with a simple formula to calculate the Gini coefficient for grouped data.

$G = C\sum_{k=1}^{K} w_k\left(1-\frac{\bar{y}_k}{\bar{y}}\right)$ where $C = \frac{2}{n(n+1)}$ and $w_k = \frac{n_k(n_k+1)}{2}$

Here n individuals (tax payers in our case) are arranged into K mutually exclusive and exhaustive income groups with n_k individuals in group $k(k = 1,2,3,\cdots,K)$. y_k is the mean income of each income group (Abounoori and McCloughan 2003).

Stata has a simple command to assign weights to observations to calculate the Gini coefficient for grouped data. So we use Stata to calculate the Gini coefficient for the entire country, for rural & urban areas and for men & women from grouped data.

Gini Decomposition: The Gini coefficient can be decomposed into between groups and within groups contributions. A broad class of inequality metrics (including Gini coefficient) can be decomposed into components reflecting only the size, mean and inequality value of each population subgroup of income source (A. F. Shorrocks 1984). When the Gini coefficient of income inequality is decomposed into between groups and within groups contributions, a residual term arises if the subgroup income ranges overlap. Lambert and Aronson (1993) provide an understanding of this residual term.

Let G be the Gini coefficient and let the population subgroups be indexed by $k = 1,2,3,\cdots,n$. The decomposition takes the form $G = G_B + \sum a_k G_k + R$ where G_B is the between groups Gini coefficient, a_k is the product of population share and income share of the subgroup k, G_k is the Gini coefficient for the income within the subgroup k and R is the residual term.

G_B is the Gini coefficient which we would get if every income in every subgroup were to be replaced by the relevant subgroup mean. R is a residual which is zero if the subgroup income ranges do not overlap. R is simultaneously both between groups and within groups term. It

measures between groups phenomenon, overlapping, that is generated by inequality within groups (see Lambert and Aronson 1993, page 1,224). R is an across groups contribution to the Gini coefficient and is positive if there is overlapping between groups. Sometimes the way R reacts to change in the subgroup characteristics is so obscure that it can cause the overall Gini value to respond in strange manners. The overlapping interpretation offers a path to the understanding of such effects. Lambert and Aronson (1993) cite a paper by Mookherjee and Shorrocks (I982) to show how change in subgroup distribution can affect the overall Gini coefficient in an unexpected fashion. In that paper (Mookherjee and A. Shorrocks 1982), the Gini coefficient is calculated for a population group containing 2 subgroups before and after making some income changes in the first subgroup. Income is redistributed from the ends to the middle in the first subgroup. The result is an increase in the first subgroup Gini coefficient, no change in between-group inequality, and yet a fall in the overall Gini coefficient (see Lambert and Aronson 1993, page 1,226). These effects can be accounted to the negative response of the residual R.

3.3.3 Bias in the Gini coefficient when using grouped data

There is one major problem in calculating the Gini coefficient from grouped data. The Gini coefficient calculated using grouped data is downward biased. One probable reason is the grouped data does not distinguish among observations within groups since each group is represented by a mean value. Van Ourti and Clarke (2011) propose 2 solutions to cope with the dependence of the Gini index on the number of groups. First approach to reduce the bias due to grouping is to fit a parametric function that satisfies the properties of a theoretical Lorenz curve. The estimated parameters can be used to calculate the Gini coefficient. The second approach is to define nonparametric bounds in the Gini index such that the lower bound corresponds to a situation where all individuals within a group are have the same mean amount of this group, while the upper bound reflects a situation where inequality is maximal in each of the groups (see Van Ourti and Clarke 2011, page 982).

3.3.4 Reconstructing income observations

Another way to remove this grouping bias is to reconstruct individual income observations from the groups. Shorrocks and Wan (2008) describe a procedure of reconstructing individual income observations from any feasible grouping pattern. They state that the characteristics of the synthetic (reconstructed) sample exactly matches that of the reported values. It is an improved method for calculating distributional indicators such as inequality values and poverty rates from grouped distribution data. They create an algorithm that allows a sample of 'income' observations to be reconstructed from any valid set of Lorenz co-ordinates.

They put 2 constraints on the synthetic sample they create from the algorithm. First they constrain the observations thus created to take positive values to ensure that the values can be computed for all commonly used inequality indices. Second they choose a sample size of 1,000 for the synthetic distribution. They observe that the scope of improvement in accuracy is very modest if one uses larger samples (see A. F. Shorrocks and Wan 2008, page 7).

We use the algorithm proposed by Shorrocks and Wan (2008) to reconstruct individual income observations of tax payers in our data and then use these observations to calculate the Gini coefficient for the entire country and for male & female and for rural & urban tax payers. To verify Shorrocks and Wan (2008) assertion that choosing a bigger sample does not improve the accuracy, we reconstruct the individual income observations for male taxpayers into synthetic samples of 1,000 and 5,000. We find the Gini coefficient of these two samples to be exactly the same. Hence we use a standard sample size of 1,000 reconstructed income observations.

3.3.5 Income share metrics

Income share metrics calculate the share in national income a sub-population accounts for. The population or the number of tax payers are split into segments such as quintiles or any other percentage and then the income share of each segment is calculated. Usually the inequality indices calculated using the income share method do not evaluate the inequalities within the segments. Despite this apparent weakness, income shares are one of the most commonly used methods in economic literature to depict inequality because they are easy to understand and makes intuitive sense. For example, Kuznets (1955) divides the US population into 5 equal quintiles and then calculates the income share of each quintile. Piketty and Saez (2003) use 10 equal deciles (10% of the population) and calculate the income share of each decile.

We use simple income share metrics in our analysis. We use reconstructed individual incomes to compute income shares since it is difficult to estimate income shares from grouped data. We compare income shares of top 1%, top 5%, top 10%, top 50% and bottom 50% across different subgroups in Norway to know more about inequalities among various population segments within Norway in 1930. Then we compare income shares of top 1%, 5% and 10% in Norway with top income shares in other Western countries and show relative levels of inequality between Norway and other Western countries in 1930.

4 Data Analysis

In this section the results from the analysis is presented. We start with the income distribution for all tax payers in the country, tax payers in rural and urban kommunes and for male & female tax payers. We then compare income distributions across major urban and rural kommunes. Afterwards we present the Gini coefficient for the entire country, for rural and urban kommunes, for men and women from grouped data. Then we compute the Gini coefficient and draw Lorenz curve from the reconstructed individual income observations created using the algorithm described by Shorrocks and Wan (2008). Finally, we use income share metrics to show the income share of top 10%, top 5%, top 1% and bottom 50% for tax payers in the entire country, tax payers in rural & urban kommunes and for male & female tax payers.

4.1 Income distribution

In case of grouped data, 2 factors determine income distribution: income group sizes and number of people in each group. Since we also have mean incomes of all income groups, it is interesting to compare mean incomes for all income groups across the country and subgroups such rural & urban tax payers and male & female tax payers. We show mean incomes of tax payers under each income group in the entire country and various subgroups in table 4.1.

Table 4.1: Mean income under different income groups, Norway 1930

Income Groups	Country	Rural	Men	Women	Income Groups	Urban
0	0	0	0	0	0	0
100- 900	561	553	588	508	100- 900	641
1,000-1,900	1,359	1,338	1,364	1,346	1,000-1,900	1,398
2,000-2,900	2,411	2,406	2,423	2,358	2,000-2,900	2,416
3,000-3,900	3,389	3,363	3,392	3,367	3,000-3,900	3,413
4,000-4,900	4,385	4,378	4,387	4,371	4,000-4,900	4,390
5,000-6,900	5,728	5,741	5,734	5,665	5,000-6,900	5,719
7,000-9,900	8,149	8,142	8,151	8,123	7,000-9,900	8,153
10,000-19,900	13,244	13,214	13,213	13,602	10,000-19,900	13,259
20,000 & above	43,281	44,439	43,269	43,410	20,000-49,900	29,276
					50,000 & above	102,566

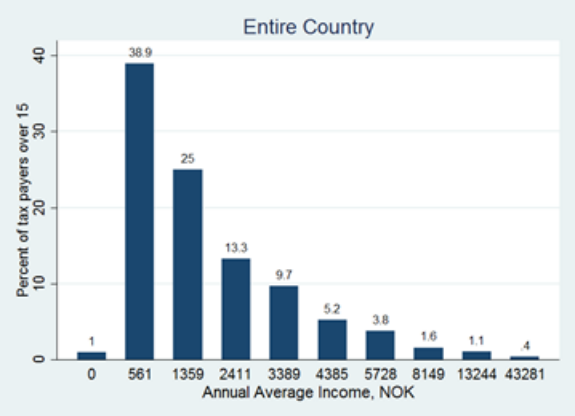
We can see that the mean incomes in each income group for all subgroups are similar with very few exceptions. The only major disconnect arises when we compare mean income of

top income group for country, rural, men, women to that of urban. The reason is obvious because these income groups do not depict the same incomes and hence are not comparable. Although the above table is informative, it does not give us the complete picture because we do not know the number of tax payers in each income group. The mean income in the lowest income group (100 - 900 NOK) in urban subgroup at 641 NOK is more than that of rural subgroup at 553 NOK but this does not imply that urban kommunes had less poor people than rural kommunes. We should compare number of people under each income group to find out more about income distribution in kommunes and entire country.

4.1.1 All tax payers

We first analyse the income distribution for all the tax payers in the country. As explained in the methodology section, we use bar charts instead of histograms to show income distribution because we want to plot mean (average) income instead of income group ranges on the x axis.

Figure 4.1: Income distribution of all tax payers, Norway 1930



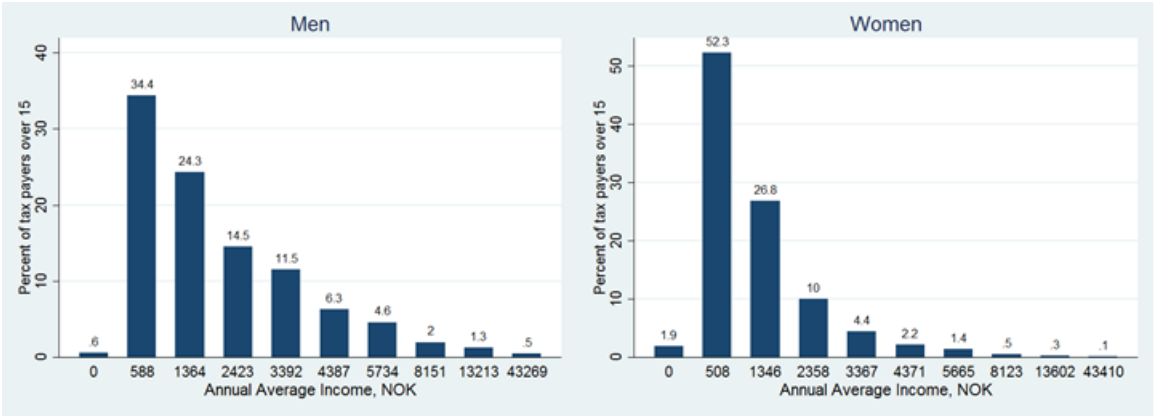
We plot income distribution of all registered tax payers in Norway in 1930 in figure 4.1. We show percentage instead of absolute number of tax payers in each income group because percentages are easier to interpret and discuss[10]. Around 39% of tax payers in the whole country had an annual income between 0 – 900 NOK with an average annual income of 561 NOK while 25% had an annual mean income of 1,359 NOK. Majority of the tax payers (78%) had annual income less than 3,000 NOK while just 3% had annual income more than 7,000 NOK & above. Rest 19% had an annual income between 3,000 to 7,000 NOK.

[10]We reproduce income distribution charts showing absolute number of tax payers for all tax payers, male & female tax payers and rural & urban tax payers in appendix A

4.1.2 Male vs. Female tax payers

We plot income distribution of male and female tax payers in figure 4.2. The number of female tax payers was much less than the number of male tax payers in Norway in 1930. One reason could be that married women were considered as a part of husband's tax unit in 1930. Another could be that a fewer women as compared to men had jobs and hence income.

Figure 4.2: Income distribution of male and female tax payers, Norway 1930



As we can see from figure 4.2, there existed a substantial difference between incomes of men and women in 1930. Mean annual income of men in the income group 0 – 900 was 588 NOK while for women it was only 508 NOK. More than half of women while only 35% of men had incomes in this range. This shows that not only that a higher percentage of women as compared to men belonged to the 0 - 900 NOK income group, average income of these women was also less than the average income of men. Although 3% of the all tax payers had annual income of 7,000 NOK & above (top 3 income groups), less than 1% of women and around 4% of men belonged to this category. So not only a higher percentage of women had low incomes, a much lower percentage of women had high income as compared to men.

4.1.3 Rural vs. Urban tax payers

The difference between rural and urban tax payers was even more stark than the difference between male and female tax payers. Norway in 1930 was predominantly a rural society with 69% of the tax payers living in rural kommunes but urban areas were more prosperous.

Figure 4.3: Income distribution of rural and urban tax payers, Norway 1930

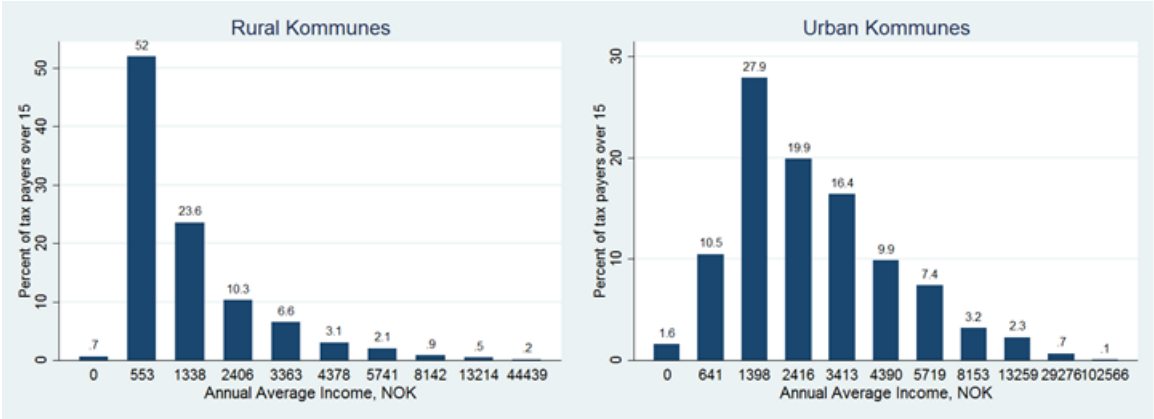


Figure 4.3 shows that a much higher percentage of rural tax payers as compared to urban ones belonged to the lower income groups. We can see that 52% of rural tax payers had income between 0 – 900 NOK while less than 11% of urban tax payers belonged to this income category. Since such a high percentage of rural tax payers had low income, it is natural that relatively low percentage of rural tax payers had higher incomes. When we consider annual income of 7,000 NOK & above, less than 2% of rural tax payers qualify as compared to more than 6% of urban tax payers. This suggests that urban kommunes were much richer than rural kommunes in 1930. We do not have occupation data for tax payers so we cannot say why urban incomes were more than rural income although it is probable that rural incomes were mostly farm based while factory wages comprised most of the urban incomes. Relative high number of top earners in urban areas indicates that more people in urban areas owned businesses or were entrepreneurs.

4.1.4 Major rural kommunes

As shown in the last subsection, rural kommunes had relatively lower share of top income earners as compared to urban kommunes. This section analyses income distribution in rural kommunes in detail by selecting 8 rural kommunes. We use 2 criteria to select kommunes: geography and number of tax payers in a kommune so that we can compare income distribution across kommunes of all sizes and from all over the country. We choose Aker, Borum, Fana and Strinda as 4 big rural kommunes and Øvre Eiker, Skjeberg, Verdal and Hadsel as 4 mid-sized rural kommunes. Aker today is a geographic area within Oslo city but was a rural kommune in the Akershus county in 1930. Even though it was a rural kommune, it had the third highest number of tax payers in the country after Oslo and Bergen. Bærum today is also a suburb of Oslo but was also a rural kommune from Akershus county in the east. Fana is an erstwhile kommune from Hordaland county in the west. Fana was merged into the Bergen city when Bergen became part of Hordaland county in 1972. Strinda is a

former kommune from the Sør-Trøndelag county in the central part of Norway. Strinda later became the part of the Trondheim city. Skjeberg from the Østfold county in the east is a former rural kommune that was merged into the Sarpsborg city while Øvre Eiker is a kommune from the Buskerd county, also located in the east. The remaining 2 kommunes, Verdal from the Nord-Trøndelag county and Hadsel from the Nordland county lie in central and northern part of the country respectively.The bigger kommunes are from the east because historically eastern part of Norway has been the most populated region and this is true even today. The 2 biggest rural kommunes (Aker and Bærum) have quite similar income distribution with less than 10% of tax payers in the 0 – 900 NOK income group. Percentage of tax payers in this income category are in the range of 25% - 33% for Fana, Strinda, Øvre Eiker and Skjeberg and this number is extremely high at 50% for Hadsel and 68% for Verdal. To simplify our analysis and make comparison easier, we categorise all income groups in 3 broad categories: low (0 – 2,900 NOK), medium (3,000 – 6,900 NOK) and high (7,000 & above NOK) and then compare percentage of tax payers in each category.

Figure 4.4: Income distribution in important rural kommunes, Norway 1930

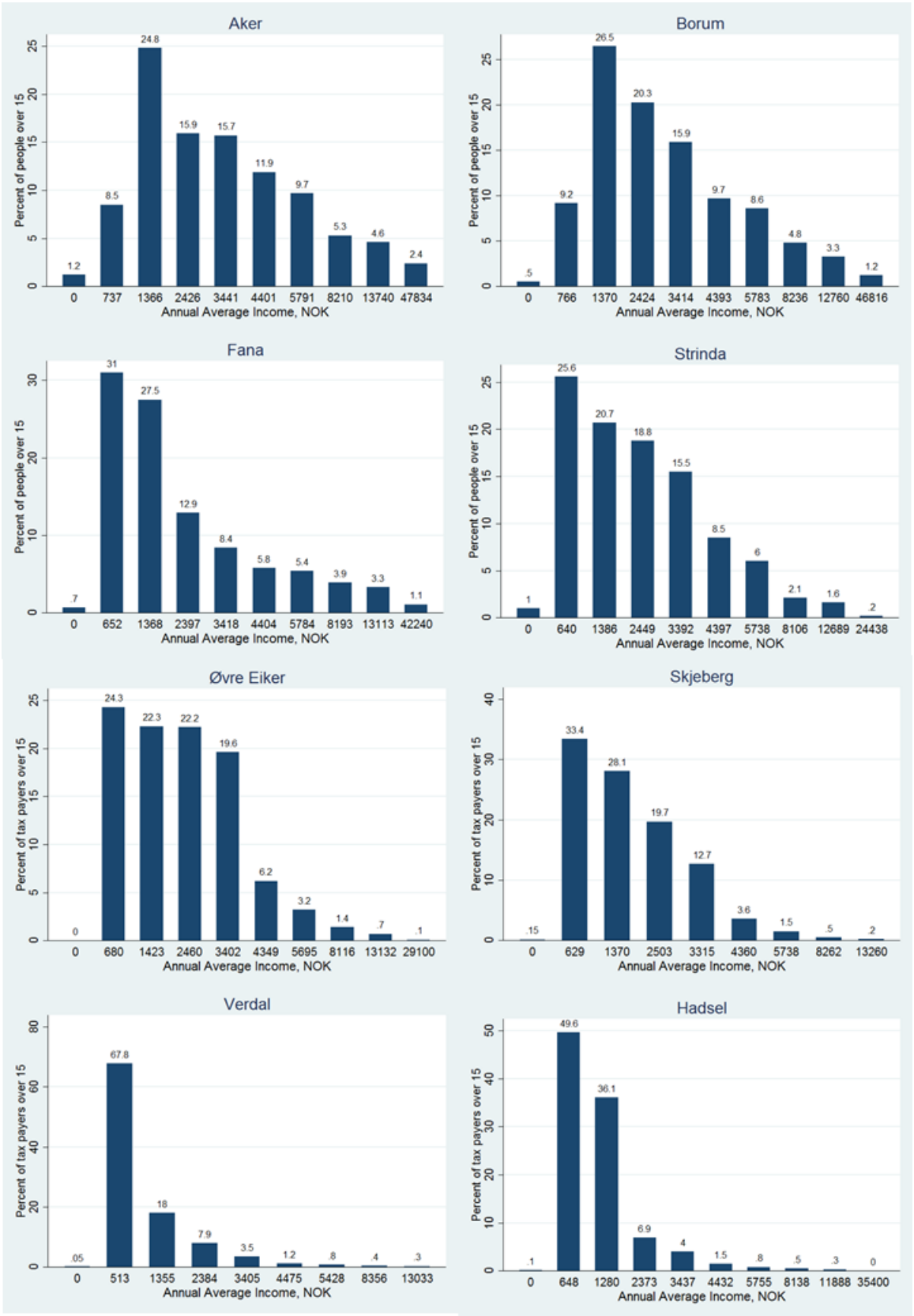


Table 4.2: Tax payers in low (0 – 2,900 NOK), medium (3,000 – 6,900 NOK) and high (7,000 & above NOK) income categories across 8 rural kommunes, Norway 1930

	Low	Medium	High
Aker	51 %	37 %	12 %
Bærum	67 %	24 %	9 %
Fana	72 %	20 %	8 %
Strinda	66 %	30 %	4 %
Øvre Eiker	69 %	29 %	2 %
Skjeberg	81 %	18 %	1 %
Verdal	93 %	6 %	1 %
Hadsel	93 %	6 %	1 %

Table 4.2 shows percentage of tax payers in low, medium and high income groups. We can see that a huge difference existed across rural kommunes in 1930. For example, both Verdal and Hadsel had 93% of tax payers in the low income group (annual income of 2,900 NOK or less) while Aker had just 51% of tax payers in this category. Aker, Bærum and Fana had relatively higher percentage of high income tax payers as compared to Øvre Eiker, Skjeberg, Verdal and Hadsel while Strinda was in between. Aker had 12%, Bærum had 9%, Fan had 8%, Strinda had 4% of total tax payers in the high income group (7,000 NOK & above) while this number was around 1.5% for all rural kommunes combined. So Aker, Bærum and Fana were on an average richer than other rural kommunes. Appendix B shows the kernel density estimation for these 8 rural kommunes. Both the bar charts and kernel density estimation have very similar results.

4.1.5 Major urban kommunes

Now we do the same analysis for urban kommunes and see how they fare against each other. We use the same criteria to select urban kommunes that we used for rural kommunes. We choose 4 big urban kommunes and 4 mid-sized urban kommunes from different regions of the country. We start with Oslo, Bergen, Nidaros (Trondheim) and Stavanger because they had the highest number of tax payers amongst urban kommunes in 1930. We select 4 mid-sized historic kommunes, Haugesund, Kongsberg, Tromsø and Tønsberg, from different parts of the country.

Nidaros that later became a part of Trondheim lies in the centre in Sør-Trøndelag county. The present day Trondheim city is made up of Nidaros and Strinda kommunes. Oslo in east and Bergen in west were the 2 biggest cities in 1930 and counties in themselves. Bergen was later merged into Hordaland county in 1973. Stavanger and Haugesund are from the

Rogaland county in the south-western part of the country. Kongsberg is from Buskerud county in the south and Tønsberg is from the Vestfold county in central Norway. The last kommune Tromsø has historically been the biggest Norwegian city in the north and lies in the Troms county.

Figure 4.4 shows the income distribution across these 8 urban kommunes. Income distribution varies much across these kommunes indicating that different parts of Norway had unequal levels of income. For example, percentage of tax payers with annual income between 0 – 900 NOK was as high as 25% in Bergen and Kongsberg, 18% in Haugesund and 14% in Tønsberg while it was less than 6% in Stavanger and less than 1% for Oslo, Nidaros and Trømso. This shows that some parts of Norway had a much higher share of low income tax payers than others.

Figure 4.5: Income distribution in important urban kommunes, Norway 1930

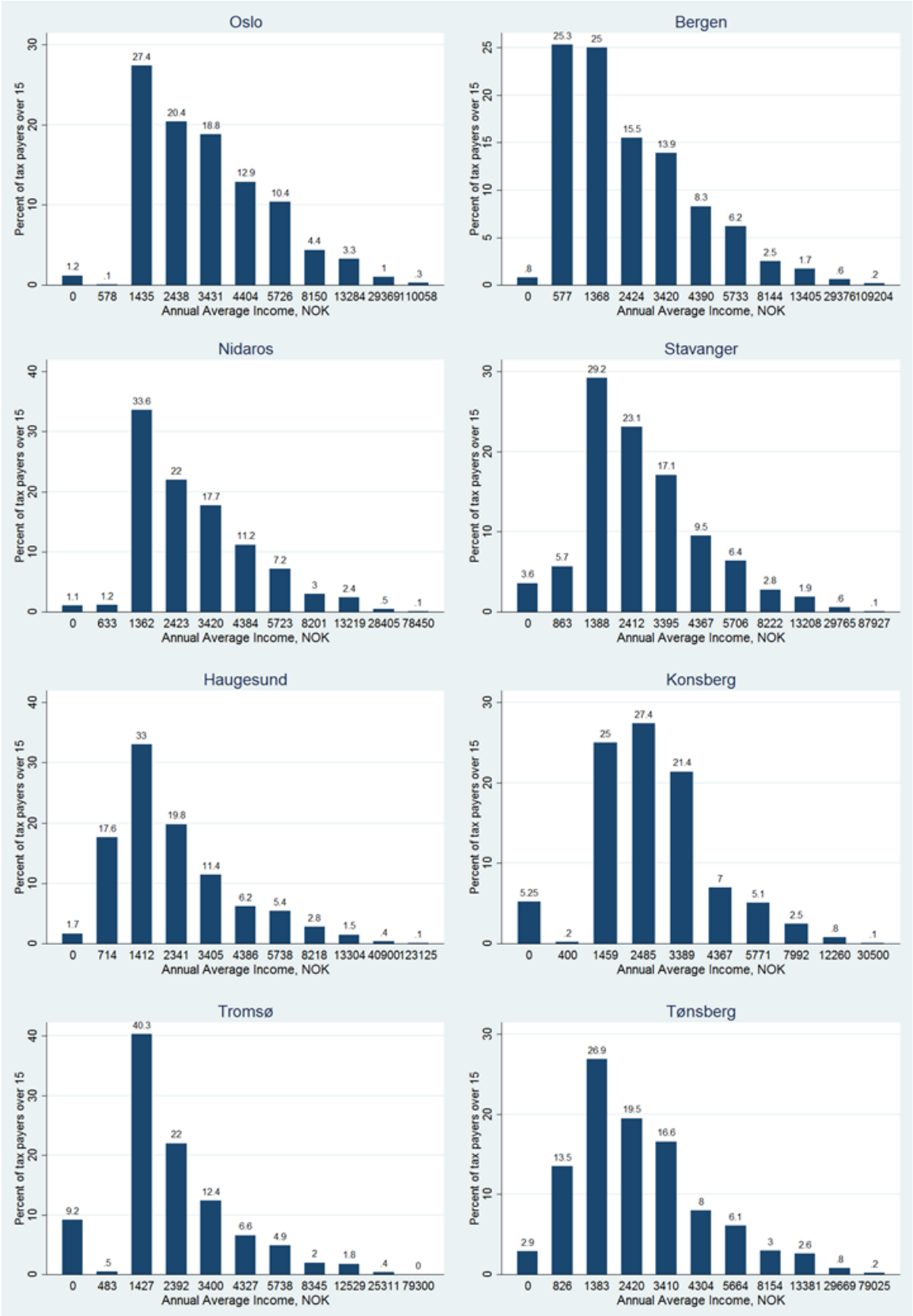


Table 4.3: Tax payers in low (0 – 2,900 NOK), medium (3,000 – 6,900 NOK) and high (7,000 & above NOK) income categories across 8 urban kommunes, Norway 1930

	Low	Medium	High
Oslo	49 %	42 %	9 %
Bergen	67 %	28 %	5 %
Nidaros	58 %	36 %	6 %
Stavanger	68 %	27 %	5 %
Haugesund	72 %	23 %	5 %
Kongsberg	84 %	15 %	1 %
Tromsø	72 %	24 %	4 %
Tønsberg	62 %	31 %	7 %

We present the percentage of tax payers in low, medium and high income groups for all 8 urban kommunes, using the same definition of low, medium and high incomes that we used for rural kommunes, in table 4.3. Oslo stands out because less than half of the tax payers had low annual income while for Nidaros this number was 58%. Percentage of tax payers in low income group was relatively high in Bergen, Stavanger, Haugesund, Tromsø, Tønsberg and extremely high for Kongsberg. Kongsberg was an extremely poor kommune with 84% of tax payers had an annual income of less than 2,900 NOK. Both Oslo and Nidaros had a greater percentage of medium tax payers as compared to other kommunes. Appendix C shows the kernel density estimation for these 8 urban kommunes. Both bar charts and kernel density estimation give similar results.

4.1.6 Neighbourhoods in Aker, Bergen and Oslo

Aker, Bergen and Oslo were the 3 biggest kommunes by the number of tax payers in 1930. Our data also has distribution of tax payers income in different neighbourhoods of these 3 kommunes. In this subsection we compare income distribution within neighbourhoods of these 3 kommunes and evaluate neighbourhoods on their share of rich and poor people. We show the percentage of tax payers in top income group (20,000 NOK & above) and bottom income group (less than 900 NOK) for these 3 kommunes in table 4.4.

Aker kommune had around 10% of tax payers with income less than 900 NOK. There was not much difference among neighbourhoods of Aker regarding concentration of low income tax payers. On the other hand, Ullern and Vestre Aker had higher concentration of high income tax payers while Østre Aker had very few high income tax payers.

All neighbourhoods in Bergen had around the same percentage of low income tax payers

while Johannes, Domkirken (present day city centre and nearby areas) had a higher percentage of high income tax payers. Sandviken and St. Olav, now considered upscale neighborhoods in Bergen, were poor neighbourhoods in 1930 with very few high income tax payers.

Income distribution within Oslo was more skewed with high income tax payers concentrated in certain neighbourhoods. Oslo was and still is segregated on social and economic lines with areas in the west known as ”West End” and the areas lying in the east known as ”East End”. Most of the rich and exclusive neighborhoods in Oslo lie in West End while East End comprises of working class and immigrant areas. Frogner and Uranineborg, both part of the West End, had extremely high concentration of rich tax payers, more than 4 times the city average. On the other hand, East End boroughs and neighbourhoods like Grønland, Kampen, Sagene had extremely low concentration of rich tax payers. This proves that, even way back in 1930, West End had much higher concentration of rich people than East End.

Table 4.4: Percentage of high (20,000 NOK & above) and low (900 NOK & below) income taxpayers in neighbourhoods of Aker, Bergen and Oslo, Norway 1930

	Low	High
Aker	**9.66 %**	**2.45 %**
Nordstrand	9.91 %	1.75 %
Ullern	9.06 %	4.62 %
Vestre Aker	8.70 %	3.69 %
Østre Aker	10.93 %	0.30 %
Bergen	**26.12 %**	**0.78 %**
Mariekirken, Korskirken, Nykirken	29.20 %	0.25 %
Sandviken, St. Olav	28.93 %	0.13 %
Ørstad, St. Markus	23.13 %	0.50 %
Johannes,Domkirken	24.05 %	1.86 %
Oslo	**1.26 %**	**1.22 %**
Vår Frelser, Trefoldighet, Jakob	1.18 %	1.04 %
Frogner og Uranienborg	2.19 %	4.93 %
Fagerborg, Gamle Aker og Markus	1.83 %	1.03 %
Sagene, Torshov, Lilleborg og Petrus	0.53 %	0.10 %
Paulus, Hauge og Mattøus	0.73 %	0.10 %
Grønland	0.81 %	0.11 %
Kampen, Vølerengen og Gamlebyen	0.97 %	0.09 %

4.2 Income inequality

The second objective of this thesis is to use various methods to calculate income inequality in Norway in 1930. As discussed in methodology section, the Gini coefficient, the Lorenz curve and income shares are some of the widely used ways to calculate the income inequality. We start our analysis with the grouped data and then move to reconstructed income observations.

4.2.1 Gini coefficients using grouped data

In this subsection, we calculate the Gini coefficient from our grouped data for all tax payers and for its subgroups: rural & urban tax payers, male & female tax payers and rural men, urban men, rural women and urban women. As explained in the methodology section, the Gini coefficient of grouped data is always downward biased so in next section we calculate the Gini coefficients from reconstructed income observations and then compare the 2 types of Gini coefficients. We use average income across different income groups for all the tax payers in the country to calculate the Gini coefficient for the entire country. To calculate the Gini coefficient for any subgroup, we use the observations from that subgroup and drop all the other observations. For example, the Gini coefficient of rural tax payers is calculated using average income of different income groups only for rural tax payers and dropping all urban tax payers from the analysis. We follow the same methodology to calculate the Gini coefficients for other subgroups.

Table 4.5: The Gini coefficients from grouped observations, Norway 1930

Subgroups	Gini
All tax payers	0.499
Rural tax payers	0.480
Urban tax payers	0.431
Male tax payers	0.494
Rural male tax payers	0.473
Urban male tax payers	0.410
Female tax payers	0.453
Rural female tax payers	0.393
Urban female tax payers	0.379

The Gini coefficient for all the tax payers in the country was 0,499 while the Gini coefficient for rural and urban tax payers was 0.480 and 0.431 respectively. This means that the inequality was higher amongst rural tax payers as compared to inequality amongst urban tax payers. The Gini coefficient of all tax payers is more than the Gini coefficients of both

rural and urban tax payers. One reason behind this could be that urban tax payers in general had higher average income than rural tax payers. Hence the difference in incomes when considering urban tax payers in isolation or rural tax payers in isolation was less than the difference in incomes for all tax payers.

4.2.2 Gini coefficient decomposition using grouped data

As explained in the methodology section, the Gini coefficient can be decomposed into within groups and between groups components and also a residual term known as overlap. We decompose Gini coefficients of all tax payers into men & women components and urban & rural components. Similarly we decompose both male and female tax payers Gini coefficients into rural & urban components.

Table 4.6: The Gini coefficients and its components, Norway 1930

All tax payers	Men	Women	Between	Overlap	Within
0.501	0.494	0.453	-0.094	0.268	0.327
All tax payers	Rural	Urban	Between	Overlap	Within
0.500	0.480	0.436	-0.175	0.441	0.234
Male tax payers	Rural	Urban	Between	Overlap	Within
0.494	0.473	0.410	-0.184	0.444	0.235
Female tax payers	Rural	Urban	Between	Overlap	Within
0.458	0.393	0.379	-0.197	0.469	0.186

The above Gini coefficients are almost identical as Gini coefficients from table 4.5. Some very minor differences exist because of the different nature of grouping of data. For example, the Gini coefficient for all tax payers in table 4.5 uses data in which all tax payers are divided into various income groups. On the other hand, the same Gini coefficient from table 4.6 uses a data in which all tax payers are first divided into men and women and then both men and women are divided into various income groups. Hence there is a minor difference between the 2 Gini coefficients (0,499 vs. 0,501). Table 4.6 shows us that the overlap component (residual term) is very high for all Gini coefficients while between groups component is negative for all Gini coefficients. Moreover, overlap term is very significant in all cases and accounts for biggest contribution to total inequality in cases when the decomposition is done between rural and urban subgroups. It means that the subgroup incomes overlap to a large extent. This is obvious because both rural and urban tax payers had largely similar incomes. We also get a high overlap component when we decompose the Gini coefficient for all tax payers into men and women components although a relatively lower overlap magnitude suggests that income varied more between men and women. Negative between groups components are trickier to explain. We can say that since the incomes were so

similar across subgroups, the interaction between subgroups is reducing the overall Gini value instead of increasing it. Surprisingly, economic literature is silent on negative between groups component of a Gini decomposition. One reason could be that usually different components of a Gini decomposition have very different incomes unlike our case when men and women or rural and urban areas have similar income.

4.2.3 Gini coefficients from the reconstructed income observations

We construct 9 datasets each consisting of 1,000 individual income observations for all tax payers, rural tax payers, urban tax payers, male tax payers, rural males, urban males, female tax payers, rural females and urban females respectively and then use these observations to calculate the Gini coefficients, Lorenz curves and income shares.

Table 4.7: The Gini coefficients from reconstructed income observation and grouped observations, Norway 1930

Subgroups	Reconstructed observations	Grouped observations
All tax payers	**0.522**	**0.499**
Rural tax payers	0.514	0.480
Urban tax payers	0.457	0.431
Male tax payers	**0.513**	**0.494**
Rural male tax payers	0.504	0.473
Urban male tax payers	0.423	0.410
Female tax payers	**0.507**	**0.453**
Rural female tax payers	0.498	0.393
Urban female tax payers	0.434	0.379

We can see from table 4.7 that Gini coefficients from the reconstructed income observations are greater than Gini coefficients from the grouped data in all cases. This is expected because Gini coefficients from the grouped observations is downward biased as shown by Van Ourti and Clarke (2011). The Gini coefficients of different subgroups from the reconstructed income observations have the same order as Gini coefficients from the grouped data. The Gini coefficient of rural tax payers is more than the Gini coefficient of urban tax payers and the Gini coefficient of male tax payers is more than the Gini coefficient of female tax payers under both methods.

The only major difference in results between the two methods is that the difference between Gini coefficients of male and female tax payers is much less in the reconstructed income observations than the grouped data. Using reconstructed income observations, we find that even though the inequality amongst female tax payers is less than the inequality amongst male tax payers, this difference in inequalities is much less than what the grouped data

depicts. One probable reason could be that the differences in incomes of individual tax payers within income groups for female tax payers is more than that of male tax payers. The Gini coefficients from grouped data ignore this within group differences in incomes. So, both female and male tax payers Gini coefficients are under reported, the magnitude is higher for female tax payers. Reconstructed income observations on the other hand considers this difference and the Gini coefficient from the reconstructed observations reflect this.

4.2.4 Lorenz curves from the reconstructed income observations

A graphical way to show income inequality is using Lorenz curves. Since the Gini coefficient is derived from the Lorenz curve, both Gini coefficient and Lorenz curve give same results. We have 9 sets of reconstructed income observations. We divide them in 3 categories and then make 3 sets of Lorenz curves. The first set of Lorenz curve shows the inequality among all tax payers and rural and urban subgroups. The second set depicts the inequality among male tax payers and subgroups while the third set shows the inequality among female tax payers and subgroups.

Figure 4.6: Lorenz curve for all tax payers, Norway 1930

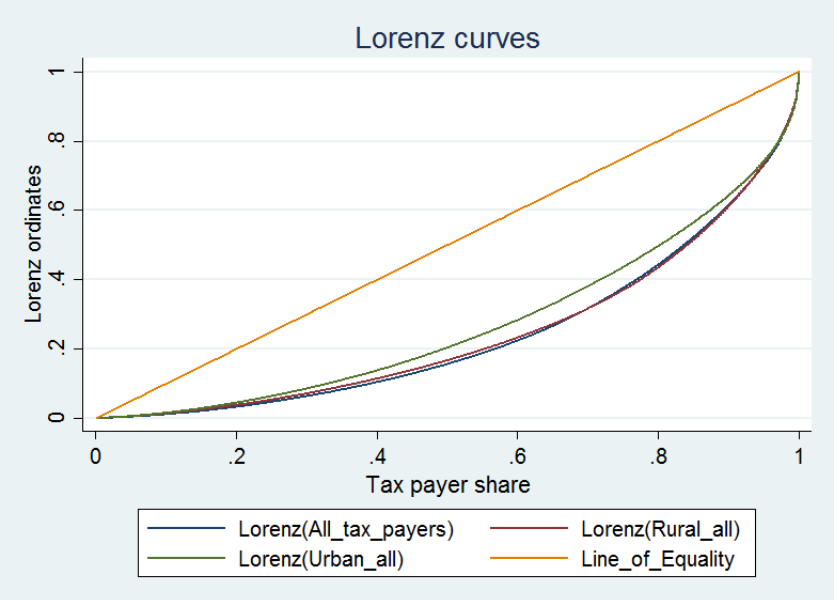


Figure 4.6 shows that the area between the Perfect equality line and the Lorenz curve for urban tax payers is much less than the area between the Perfect equality line and the Lorenz curve for all tax payers and the area between the Perfect equality line and the Lorenz curve for rural tax payers. According to the definition of Lorenz curve, we can say that urban tax payers have lower level of inequality when compared to all tax payers and rural tax payers.

The Gini coefficients[11] also give us the same inference.

Figure 4.7: Lorenz curve for male tax payers, Norway 1930

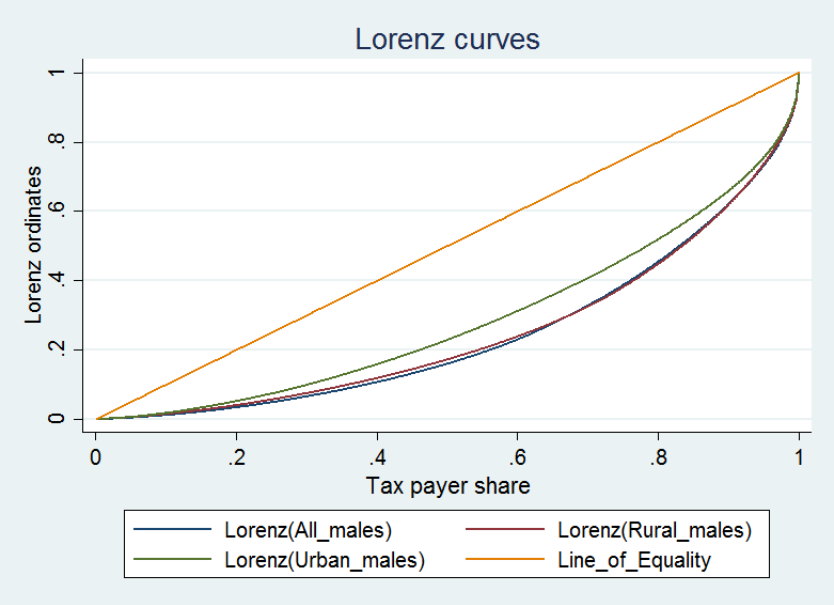
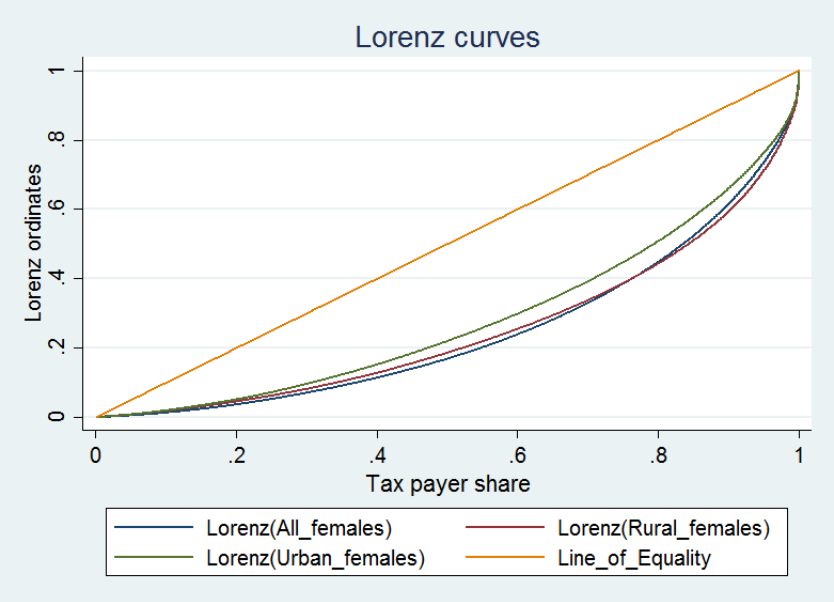


Figure 4.8: Lorenz curve for female tax payers, Norway 1930



We reach the same conclusion for male and female tax payers from the Lorenz curves in figures 4.7 and 4.8: Both urban men and urban women had lower level of inequality as compared to rural men and all men and rural women and all women. This is also evident from the Gini coefficients of urban men and urban women.

[11] see table 4.7

4.2.5 Top income share for all tax payers and various subgroups

We use the same set of reconstructed income observations that we use to compute Gini coefficients to estimate top 1%, 5%, 10%, 50% and bottom 50% income shares. To estimate top income shares of a subgroup, we use observations only from that subgroup and not observations from other subgroups. For example, we use incomes observations from male tax payers subgroup to calculate top income share of of male tax payers subgroup. We do not consider income observations of other subgroups while computing top income share of male tax payers subgroup. Similarly, top income shares of urban female tax payers consider income observations of just urban female tax payers subgroup and no other observations. So when we say income share of top 1% of uban male tax payers was 12.7%, we mean top 1% urban male tax payers had 12.7% share in total income accounted to urban male tax payers and not to the entire population. We present our top income share results in the discussion section.

5 Discussion

The objective of this thesis is to study income distribution and income inequality in Norway in 1930. So far in this report, we have documented income distribution and income inequality in Norway using different metrics. This section builds upon the previous analysis by giving explicit and clearer answer to the question we set out to answer in the beginning to this thesis. The question is:

How did income distribution and inequality look like in Norway in 1930?

We begin our answer by analysing income distribution in Norway in 1930 in detail. We have already shown income distribution using bar charts for the entire country, men & women, rural & urban areas and important kommunes. Although absolute numbers in isolation do not tell us much, when we compare these numbers with each other a picture of Norwegian society in 1930 emerges. For example, when we say 34% of male tax payers were in the lowest income group with a mean annual income of 588 NOK while 52% of female tax payers belonged to the lowest income group with mean annual income of 508 NOK, we see a society where men were richer than women. In this section, we build upon such comparisons to debate more about income distribution in Norway in 1930.

We use 2 indicators, Gini coefficients and income shares that we computed in the data analysis section, to compare inequality across different population subgroups in Norway. We also compare our results with other inequality research in Norway from the same period and discuss variances, if any. Finally, we compare Norway with other Western countries and comment on relative levels of inequality between Norway and other countries.

5.1 Income distribution in Norway in 1930

We started our analysis of income distribution[12] by dividing all the tax payers in 4 subgroups: men, women, rural and urban. Now we subdivide male tax payers into rural and urban and female tax payers into rural and urban. We have already analysed income distribution of the original 4 subgroups in detail. Now we summarise the income distribution of these new subgroups by dividing income in each subgroup in 3 categories: low (0 – 2,900 NOK), medium (3,000 – 6,900 NOK) and high (7,000 & above NOK). We use this distribution to compare tax payer concentration in each category across subgroups.

Table 5.1: Tax payers in low (0 – 2,900 NOK), medium (3,000 – 6,900 NOK) and high (7,000 & above NOK) income categories, Norway 1930

	Low	Medium	High
All tax payers	78.2 %	18.7 %	3.1 %
Rural tax payers	86.6 %	11.8 %	1.6 %
Urban tax payers	63.2 %	33.7 %	6.3 %
Male tax payers	73.8 %	22.4 %	3.8 %
Rural male tax payers	84.0 %	14.1 %	1.9 %
Urban male tax payers	**48.2 %**	**43.3 %**	**8.5 %**
Female tax payers	91.1 %	8.0 %	0.9 %
Rural female tax payers	**96.1 %**	**3.5 %**	**0.4 %**
Urban female tax payers	87.4 %	10.9 %	1.7 %

We can see from table 5.1 that a disproportionately larger number of high income tax payers lived in urban areas. Men had much greater percentage of high income tax payers as compared to women. Urban areas and men had a higher share of medium income tax payers as well when compared to rural areas and women respectively. We can that huge differences existed among subgroups. Urban men as a subgroup stand out because of extremely low concentration of low income tax payers and high concentration of medium and high income tax payers. For example, 96.1% of rural women were in the low income group while just 48.2% of urban men had low income. This difference is staggering because this shows us how poor rural women were as compared to urban men. A big difference also existed between incomes of men living in urban and rural areas. Eighty four percent of rural men and 91 % of urban women were from low income group. Hence, income distribution in rural men was more similar to urban women and rural women than to urban men. Differences between rural and urban income distribution were greater than the differences between men and women income distribution.

[12] see section 4.1

We can use this concentration of low, medium and high tax payers to rank different subgroups. Men as a subgroup were on an average richer than women and urban areas were richer than rural areas. Urban men were the richest subgroup, followed by rural men and urban women while rural women were the poorest subgroup. The difference between men and women was much more in urban areas than in rural areas.

In the next section we build upon this analysis by comparing top income shares across these subgroups. We can compare a specific top percentage of population for all subgroups through top income shares analysis. For example, we can compare top 1% of urban men against top 1% of rural women using top income shares.

5.2 Top income shares in Norway in 1930

We discuss top income shares first and then continue to Gini coefficients because income share distinguishes between different quantiles of the population while Gini coefficient is a more wholesome indicator that shows degree of overall inequality for the entire population. Even though income shares do not condense inequality within a group to one number like Gini coefficient does, they are easy to understand and suitable for comparison purpose.

Table 5.2: Income shares, Norway 1930

	Bottom 50%	Top 50%	Top 10%	Top 5%	Top 1%
All tax payers	15.66 %	84.34%	38.46 %	26.93 %	12.55 %
Rural tax payers	16.72 %	83.28%	38.85 %	26.65 %	11.90 %
Urban tax payers	20.41 %	79.59%	35.68 %	25.87 %	12.68 %
Male tax payers	16.03 %	83.97%	37.90 %	26.79 %	12.60 %
Rural male tax payers	17.25 %	82.75%	38.02 %	26.40 %	12.40 %
Urban male tax payers	22.93 %	77.07%	34.09 %	24.49 %	11.48 %
Female tax payers	16.91 %	83.09%	38.42 %	26.52 %	12.43 %
Rural female tax payers	18.63 %	81.37%	40.61 %	28.90 %	13.70 %
Urban female tax payers	22.01 %	77.99%	33.95 %	23.75 %	11.91 %

We create table 5.2 using the same reconstructed income observations that we use to estimate Gini coefficients of various subgroups. We can see that the income shares of top 1% varied between 11.48% for urban men and 13.70% for rural women[13]. This is interesting because we know that urban men were the richest subgroup while rural women were the poorest. Now we see that top 1% income share was lowest in urban men and highest in rural women. Does this mean that urban men were the richest as well as the most equal subgroup while the reverse was true for rural women? Although top income shares do tell us that income was less concentrated in top 1% of urban men as compared to top 1% of rural women, we must compare Gini coefficients to answer this question about overall inequality between urban men and urban women.

When we consider top 5% income share, urban women had the lowest income share at 23.75% while urban men came close second at 24.49%. Rural women had the highest top 5% income share at 28.90% followed by all tax payers at 26.93%. Urban women had the lowest top 10% income share at 33.95% and were closely followed by urban men at 34.09%. Rural women had the highest top 10% income share at 40.61% followed by all rural tax payers at 38.85%. So rural women had highest top 1%, 5% and 10% income share amongst all subgroups.

If we consider income share of bottom 50%, urban men had the highest share at 22.93%

[13] See section 4.2.5 to know about top income share definition within subgroups.

followed by urban women at 22.01%. Bottom 50% income share of all tax payers was lowest at 15.66% and male tax payers and tax payers from rural areas had second and third lowest bottom 50% income share with 16.03% and 16.72% respectively. Although top and bottom shares varied a lot between rural and urban tax payers, they were quite similar for male and female tax payers. This proves that there was a bigger difference between rural and urban areas than between men and women.

The next question is how does our results compare with results of other economists? We have already shown in table 1.1[14] the comparison between top income shares in Norway and other Western countries in the period before World War II. The Norwegian top income share in the table 1.1 is from Aaberge and Atkinson (2010) who use Norwegian municipal and central government income tax records to construct an income series dating from 1875 till 2006. They calculate top income share of top 1%, 5% and 10% to be 12.57%, 28.25% and 41.32% respectively in the year 1929 (see table 9.1 Aaberge and Atkinson 2010, page 454). We estimate top income share of top 1%, 5% and 10% to be 12.55%, 26.93% and 38.46% respectively in the year 1930. Although top 1% income share is exactly the same, we get different result for top 5% and top 10% income shares. The reason is Aaberge and Atkinson (2010) consider data associated with the population census in 1929. Aaberge and Atkinson (2010) use tabulations of the distribution of income as assessed for tax purposes using the number of income recipients and total amount of income by ranges of assessed income to compute income shares in 1929 while we use an extensive unpublished data set from Statistics Norway's archive covering the whole population on pretax income from the tax registries from each municipality in Norway in 1930. This explains why there is a slight difference in top income shares because we use kommune level data and then aggregate it while Aaberge and Atkinson (2010) use country level data.

To compare top income share in Norway with other Western countries in 1930[15], we reproduce table 1.1 but use our estimates for Norway instead of Aaberge and Atkinson (2010). Unfortunately we do not have bottom 50% and top 50% income share for other countries so we can not compare bottom 50% and top 50% income shares.

[14] See page 5.

[15] UK estimates are from 1937 and German estimates are from 1928

Table 5.3: Top income shares in select Western countries and Norway

	10 %	5 %	1 %
USA (1930)	43.07 %	31.18 %	16.42 %
UK (1937)	38.37 %	29.75 %	16.98 %
Netherlands (1930)	43.02 %	32.41 %	17.15 %
Canada (1930)	NA	32.74 %	16.10 %
France (1930)	41.08 %	30.14 %	15.31 %
Germany (1928)	32.20 %	22.60 %	11.20 %
Norway (1930)	38.46 %	26.39 %	12.55 %

Source: (Piketty and Saez 2003), (Atkinson and Salverda 2005), (Saez and Veall 2005), (Piketty 2007), (Dell 2007), (Norway Census 1930)

Top 1% income share in Norway was the second lowest at 12.55% after Germany at 11.20%. All other countries had much higher top 1% income share with Netherlands as high as 17.15%, USA at 16.42%, Canada at 16.10% and France at 15.31%. We do not have UK estimates for 1930 but top 1% income share in UK was 16.98% in 1937. So Norway had one of the lowest top 1% income share in 1930 and the share was much lower as compared to all major Western countries except Germany.

Top 5% income share in Norway was again less as compared to other countries except Germany. Top 5% income share was highest for Canada at 32.74%, followed by Netherlands, USA and France at 32.41%, 31.18% and 30.14% respectively while it was much lower at 26.39 % for Norway. This trend was same for top 10% income share although the differences between Norway and other Western countries except Germany decrease when we compare top 10% income share. Norway's top 10% income share was 38.46% and it was much lower at 32.20% for Germany. On the contrary, top 10% income share was highest for USA at 43.07% followed by Netherlands and France at 43.02% and 41.08% respectively. We can see that Norway and Germany were the countries with lowest top income shares in 1930[16] while USA, UK, Netherlands and France had much higher top income shares. We observe some interesting trends while comparing top incomes shares across countries. Norway, for example, had much lower top 1% income share compared to USA, Netherlands, UK and France but the differences in top 5% and 10% income shares were not so high. Does that mean differences in income shares of P 90-95 and P 95-99 were much less than difference in income share of P 99-100? To answer this question, we calculate "shares within groups" of the top decile for all these countries. P 99-100 is same as top 1 percentile of population while P 95-99 is population between 95th and 99th percentile and P 90-95 is population

[16] German estimates are from 1928.

between 90th and 95th percentile.

Table 5.4: Income shares of groups within the top decile in select Western countries and Norway

	P 90-95	P 95-99	P 99-100
USA (1930)	11.89%	14.76%	16.42%
UK (1937)	8.62%	12.77%	16.98%
Netherlands (1930)	10.61%	15.26%	17.15%
Canada (1930)	NA	16.64%	16.10%
France (1930)	10.94%	14.83%	15.31%
Germany (1928)	9.60%	11.40%	11.20%
Norway (1930)	12.07%	13.84%	12.55%

Source: (Piketty and Saez 2003), (Atkinson and Salverda 2005), (Saez and Veall 2005), (Piketty 2007), (Dell 2007), (Norway Census 1930)

Dividing top decile in different percentiles gives us interesting results. Although P 99-100 (top 1 %) income shares of Norway is lowest after Germany, P 90-95 income share in Norway is the highest. P 95-99 income share of Norway is lower than all countries except UK, where estimates are from 1937, and Germany. In UK P 99-100 income share was twice that of P 90-95 while in Norway income shares of P 99-100 and P 90-95 were almost equal. There is an obvious pattern for USA, UK, Netherlands and France where P 99-100 had the highest income share, followed by income share of P 95-99 while the income share of P 90-95 was the least. For Norway and Germany, income shares of all three groups within top decile were much more equally distributed. These shares within groups tell us that top decile income in Norway was much more equally divided among P 90-95, P 95-99 and P 99-100 as compared to countries like USA, UK, Netherlands and France where P 99-100 had a relatively higher income share. We can say that Norway's income distribution within the top decile was quite similar to Germany and very different from USA, UK, Netherlands and France. It is fair to say that top 1% contributed much less to overall income inequality in Norway as compared to most other Western countries. In the next section, we use Gini coefficient estimates to compare inequality among different subgroups in Norway. We also compare our estimates of Norway with estimates of different economists for Norway and other Western countries.

5.3 Gini coefficient in Norway in 1930

We use Gini coefficients from reconstructed income observations for our analysis because they are more accurate than grouped Gini coefficients[17]. We compare Gini coefficients of the same subgroups for whom we compared top and bottom income shares in the previous section.

Table 5.5: The Gini coefficients from reconstructed income observation, Norway 1930

Subgroups	Reconstructed observations
All tax payers	**0.522**
Rural tax payers	0.514
Urban tax payers	0.457
Male tax payers	**0.513**
Rural male tax payers	0.504
Urban males tax payers	0.423
Female tax payers	**0.507**
Rural female tax payers	0.498
Urban female tax payers	0.434

We can see from table 5.5 that the Gini coefficient for the entire country was 0.522 in 1930, much less than 0.252, the Gini coefficient of Norway in 2013 (OECD 2013). This shows that Norway in 1930 was a very unequal society and it gradually became more equal because of industrialisation and development of welfare state model that started mid 1930s. Among various subgroups, Gini coefficient of urban tax payers at 0.457 was much lower than Gini coefficient of rural tax payers at 0.514. This indicates that inequality was less among urban tax payers than rural tax payers. Outcome from top income share analysis[18] is also same. The Gini coefficients of men and women were almost equal with men having slightly higher Gini coefficient at 0.513 than women at 0.507. Top and bottom income shares of men and women were also very similar and close Gini coefficients of men and women verify that men and women subgroups had similar level of inequality.

Within men, urban men had a much lower Gini coefficient of 0.423 than Gini coefficient of 0.504 for rural men. Amongst women subgroups, Gini coefficient of urban women at 0.434 was less than Gini coefficient of rural women at 0.498. While comparing top and bottom income shares of these subgroups in the last section, we wanted to know whether urban men were the most equal subgroup and rural women were most unequal subgroup. Comparing Gini coefficients give us the answer. Urban men were indeed the most equal subgroup.

[17] See section 3.3.3 for the reasons behind using Gini coefficients from reconstructed income observations.

[18] See table 4.9 for top income share

Although rural women appeared most unequal subgroup when comparing top and bottom income shares, Gini coefficients tell us that rural men were the most unequal subgroup and rural women came close second.

Paukert (1973) computes Gini coefficients for men in 8 Norwegian cities in 1930[19]. All 8 cities in Paukert's sample are from 2 counties: Østfold in the east and Vest-Agder in the south. Gini coefficients of men in these 8 cities varied between 0.365 and 0.432. We estimate the Gini coefficient for urban men to be 0,423 in 1930. We consider all the urban areas in the country while Paukert's 8 cities are from 2 counties. Even then our estimates are not very different from each other. Aaberge, Atkinson and Modalsli (2016) explore a new series on the distribution of income in Norway as a whole spanning the period from 1875 to 2013. They use municipal and central government tax assessments in Norway to estimate the Gini coefficients in Norway from 1875 to 2013. Since the data is not available for all the years, they make assumptions and create an upper and lower bound for Gini coefficients. They choose assumptions that lead to higher inequality for the upper bound and lower inequality for the lower bound and thus they are able to efficiently bracket the true Gini coefficient that they would get if they have complete information (see Aaberge, Atkinson, and Modalsli 2016, for the methodology behind creating upper and lower bounds of Gini coefficients). They find that gross income inequality rose in Norway in the period between 1923 and 1939, largely due to increase in inequality among the upper half of the population. They estimate that the Gini coefficient in 1930 was between 0.577 and 0.567 (see Aaberge, Atkinson, and Modalsli 2016, table A1, page 40). Aaberge, Atkinson and Modalsli (2016) combine the tax data with independent estimates of the total number of tax units and total of household income to arrive at an estimate of income inequality across the entire population. This explains why their results are different from ours because we consider only the registered tax payers. Jeanette Fjære (2014) analyses income development in 16 Norwegian municipalities including 5 largest cities: Oslo, Bergen, Trondheim, Stavanger and Kristiansand. For the period between 1884 and 1966, she uses income data from municipal and state tax statistics to estimate income inequality for these 16 municipalities. This data is in a tabular form showing aggregated figures and number of tax payers in each municipality so she computes income inequality by dividing 15-year-old and above population in each municipality into 4 income groups and creating 3 interior points on the Lorenz curve. She estimates mean income inequality[20] for urban municipalities (5 big cities) in 1930 and finds it to be 0.431 (see table 5.1.1 Fjære 2014, page 42). Her estimate is close to our Gini coefficient of 0.457 for urban kommunes even though she computes Gini coefficient for just 5 cities, uses an entirely different methodology and an aggregated country level data instead of grouped kommune

[19] See table 2.2

[20] Similar to the Gini coefficient.

level data like us.

When comparing Gini coefficient in Norway with Gini coefficient in USA and World, we get interesting results. Paukert mentions that the Gini coefficient in USA in 1935-36 was 0.47 (Paukert 1973). We do not have Gini coefficient of USA in 1930 but assuming Gini coefficient in 1930 was in similar range as Gini coefficient in 1935-36, we see that Gini coefficient in USA was less than Gini coefficient of 0.522 in Norway. This means that overall inequality in USA was less than in Norway although income shares of top 1%, 5% and 10% in USA were higher than in Norway. Bourguignon and Morrisson (2002) analyse inequality among world citizens during 1820 – 1992 [21]. They estimate the Gini coefficient of the world in 1929 to be around 0.616 and declare that it kept increasing over the 20th century. This proves that inequality in Norway was much less than the average inequality in world and while the inequality in Norway declined after 1930, it increased for the rest of the world.

[21] See section 2.2 for the data they use.

6 Conclusion

We investigated income distribution and inequality for all registered tax payers in Norway in 1930 in this thesis. We used an extensive unpublished data set from Statistics Norway's archive covering the whole population on pretax income from the tax registries from each municipality in Norway (Norway Census 1930). No one in our knowledge has used this data to study income distribution and inequality in Norway in 1930. We had access to kommune level data for both men and women which enabled us to study income distribution across major urban and rural kommunes as well as for key subgroups such as urban men, urban women, rural men and rural women. We compared income distribution across different parts in Norway by studying select important kommunes from all regions of the country.

We started by reviewing important literature on inequality for major Western countries in the periods before and after the World War II. We compared top income shares in Norway with other Western countries and discovered that top income share in Norway was less than most Western countries except Germany before the World War II. Then we described our data and discussed methodology to measure income distribution and inequality. Our data pertains to registered tax payers from 742 kommunes across 20 counties in Norway. We subdivided the data into different subgroups and then estimated income distribution and inequality for all tax payers and subgroups.

In the next part of thesis, we used our findings to answer the question that we raised in the beginning of the thesis.

How did income distribution and inequality look like in Norway in 1930?

We found that in the entire country, 78% of tax payers had annual income less than 2,900 NOK while only 3% had an annual income of more than 7,000 NOK. Women as a subgroup were poorer than men and 91% of women had annual income less than 2,900 NOK as compared to 74% men. Urban areas were on average richer than rural areas and just 63% of urban tax payers had annual income less than 2,900 NOK as compared to 87% of rural tax payers. Amongst the 4 subgroups (urban men, rural men, urban women and rural women), urban men were the richest subgroup with just 48% of the tax payers having annual income less than 2,900 NOK. Rural men with 84%, urban women with 87% were second and third richest subgroups while rural women were the poorest subgroup with 96% having annual income of less than 2,900 NOK. Although urban areas were richer than rural areas, not all urban kommunes were equally rich. Out of the major urban kommunes, Oslo in east was the richest while Kongsberg in south was the poorest. Just 49% of tax payers in Oslo had an annual income of less than 2,900 NOK as compared to 63% for all urban kommunes

and 84% for Kongsberg. Bergen in west with 67% of tax payers having annual income of less than 2,900 NOK was closer to urban kommune average while Tromsø in north and Haugesund in west were relatively poorer urban kommunes with 72% of their tax payers having annual income of less than 2,900 NOK. Rural kommunes also varied a lot in income distribution with Aker, a kommune from east and now part of Oslo urban area, having 51% tax payers with annual income of less than 2,900 NOK against 87% for all rural kommunes combined. Verdal and Hadsel, 2 kommunes from west and north respectively with 93% of tax payers having annual income of less than 2,900 NOK, were among the poorest rural kommunes. This shows that Oslo and its neighbouring areas in the east were richer while areas in north and west were relatively poor. Difference in incomes also existed within big kommunes and certain neighborhoods were richer than other neighbourhoods in Oslo, Bergen and Aker.

After analysing income distribution, we compared inequality among different subgroups and inequality in entire Norway with other countries. We compared top and bottom income shares among different subgroups in Norway and found that urban men had lowest top 10% income share and highest bottom 50% income share while rural women had highest top 10% income share and lowest bottom 50% income share. Urban tax payers as one group had a higher bottom 50% and lower top 10% income share as compared to rural tax payers. On the other hand, top 10% and bottom 50% income share of male and female tax payers were very similar. When comparing top income share in Norway with other countries, we found that top 1% income share in Norway was much less than other Western countries such as USA, UK, Netherlands and France. Although top 10% income share was also less in Norway as compared to these countries, most of this difference was due to the big difference in top 1%. The differences among income shares of P 90-95 and P 95-99 percentiles in Norway and other countries were much less than differences among top 1 percentile. This proves that top 1 percent contributed much less towards income inequality in Norway compared to other Western countries in 1930.

We also used Gini coefficient to compare population subgroups in Norway and Norway as a whole against other countries. Although we computed Gini coefficients from both grouped data and reconstructed income observations, we used Gini coefficients from reconstructed income observations for our final analysis because they are more accurate. The Gini coefficients reconfirmed the inferences we arrived at from income share analysis. The Gini coefficient of urban men was the lowest while that of rural men was the highest. The Gini coefficient of rural women was slightly less than that of rural men and much more than Gini coefficients of urban women and urban men. The Gini coefficients of men and women were very close to each other while the Gini coefficient of urban areas was much lower than that

of rural areas.

We summarize our results by saying that urban areas were richer and more equal than rural areas and although men were richer than women, both men and women had similar levels of inequality. Urban men were the richest and the most equal subgroup while rural women were the poorest subgroup and rural men were the most unequal. The Gini coefficient of Norway in 1930 was 0.522, much more than the Gini coefficient of Norway in 2013 that stood at 0.252. Finally, the Gini coefficient of Norway in 1930 was in the same range as the Gini coefficient of USA and much less than the Gini coefficient of the world.

References

Article Sources

Aaberge, Rolf (2007). "Gini's nuclear family". In: *The Journal of Economic Inequality* 5.3, pp. 305–322.

Aaberge, Rolf and Anthony B Atkinson (2010). "Top incomes in Norway". In: *Top Incomes: A Global Perspective* 2.

Aaberge, Rolf, Anders Björklund, et al. (2002). "Income inequality and income mobility in the Scandinavian countries compared to the United States". In: *Review of Income and Wealth* 48.4, pp. 443–469.

Abounoori, Esmaiel and Patrick McCloughan (2003). "A simple way to calculate the Gini Coefficient for grouped as well as ungrouped data". In: *Applied Economics Letters* 10.8, pp. 505–509.

Atkinson, Anthony B, Thomas Piketty, and Emmanuel Saez (2011). "Top incomes in the long run of history". In: *Journal of economic literature* 49.1, pp. 3–71.

Atkinson, Anthony B and Wiemer Salverda (2005). "Top Incomes in the Netherlands and the United Kingdom over the 20th Century". In: *Journal of the European Economic Association* 3.4, pp. 883–913.

Bellù, Lorenzo Giovanni and Paolo Liberati (2006). "Inequality analysis: The gini index". In: *FAO, EASYPol Module* 40.

Bourguignon, Francois and Christian Morrisson (2002). "Inequality among world citizens: 1820–1992". In: *The American Economic Review* 92.4, pp. 727–744.

Dell, Fabien (2007). "Top incomes in Germany throughout the twentieth century: 1891–1998". In: *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English Speaking Countries*, pp. 365–425.

Fjære, Jeanette Strøm (2014). "Income Development in Norwegian Municipalities A Descriptive Analysis of 16 Norwegian Municipalities Over 150 Years". In:

Kuznets, Simon (1955). "Economic growth and income inequality". In: *The American economic review* 45.1, pp. 1–28.

Lambert, Peter J and J Richard Aronson (1993). "Inequality decomposition analysis and the Gini coefficient revisited". In: *The Economic Journal* 103.420, pp. 1221–1227.

Milanovic, Branko (1994). "The Gini-Type Functions: An Alternative Derivation". In: *Bulletin of Economic Research* 46.1, pp. 81–90.

Mookherjee, Dilip and Anthony Shorrocks (1982). "A decomposition analysis of the trend in UK income inequality". In: *The Economic Journal* 92.368, pp. 886–902.

Paukert, Felix (1973). "Income distribution at different levels of development: A survey of evidence". In: *Int'l Lab. Rev.* 108, p. 97.

Piketty, Thomas and Emmanuel Saez (2003). "Vol. CXVIII February 2003 Issue". In: *The Quarterly Journal of Economics* 1.

Saez, Emmanuel and Michael R Veall (2005). "The evolution of high incomes in Northern America: lessons from Canadian evidence". In: *The American Economic Review* 95.3, pp. 831–849.

Sejersted, Francis (1992). "A theory of economic and technological development in Norway in the nineteenth century". In: *Scandinavian Economic History Review* 40.1, pp. 40–75.

Shorrocks, Anthony F (1984). "Inequality decomposition by population subgroups". In: *Econometrica: Journal of the Econometric Society*, pp. 1369–1385.

Van Ourti, Tom and Philip Clarke (2011). "A simple correction to remove the bias of the Gini coefficient due to grouping". In: *Review of Economics and Statistics* 93.3, pp. 982–994.

Book Sources

Piketty, Thomas (2007). *Income, Wage, and Wealth Inequality in France, 1901–98*. United States: Oxford University Press Inc., New York.

– (2014). *Capital in the twenty-first century*. Belknap Press.

Soltow, Lee (1965). *Toward income equality in Norway*. Madison, U. of Wisconsin P.

Data Sources

Norway Census (1930). *Municipality level pretax income data from Municipality Tax registries*. Archives, Stats Norway Kongsberg.

Online Sources

OECD (2013). *Income inequality Gini coefficient*. URL: `https://data.oecd.org/inequality/income-inequality.htm` (visited on 12/12/2016).

Technical Report Sources

Aaberge, Rolf, Anthony B Atkinson, Jørgen Modalsli, et al. (2016). *On the measurement of long-run income inequality. Empirical evidence from Norway, 1875-2013*. Tech. rep.

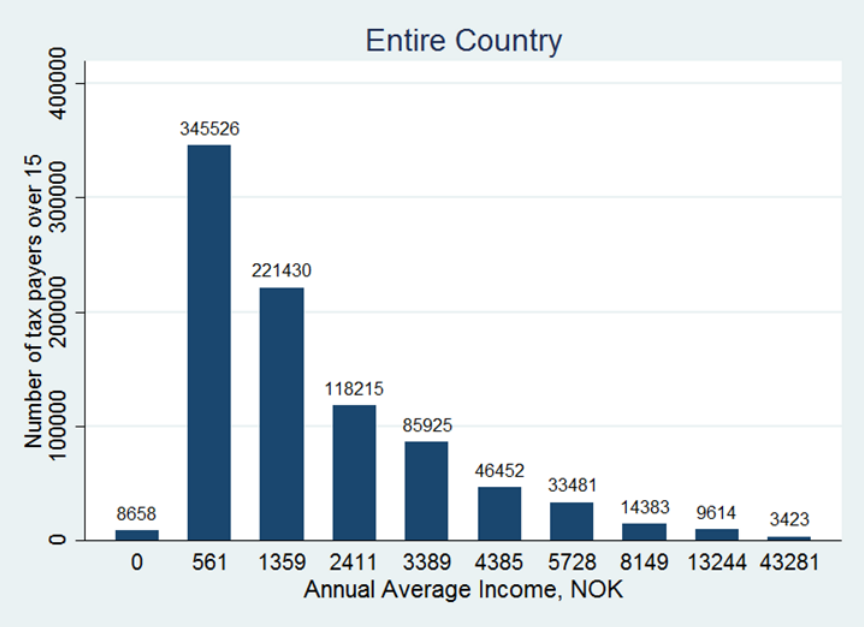
Shorrocks, Anthony F, Guanghua Wan, et al. (2008). *Ungrouping income distributions: Synthesising samples for inequality and poverty analysis*. Tech. rep. UNU-WIDER, United Nations University (UNU).

Appendices

A Appendix: Income distribution in Norway in 1930

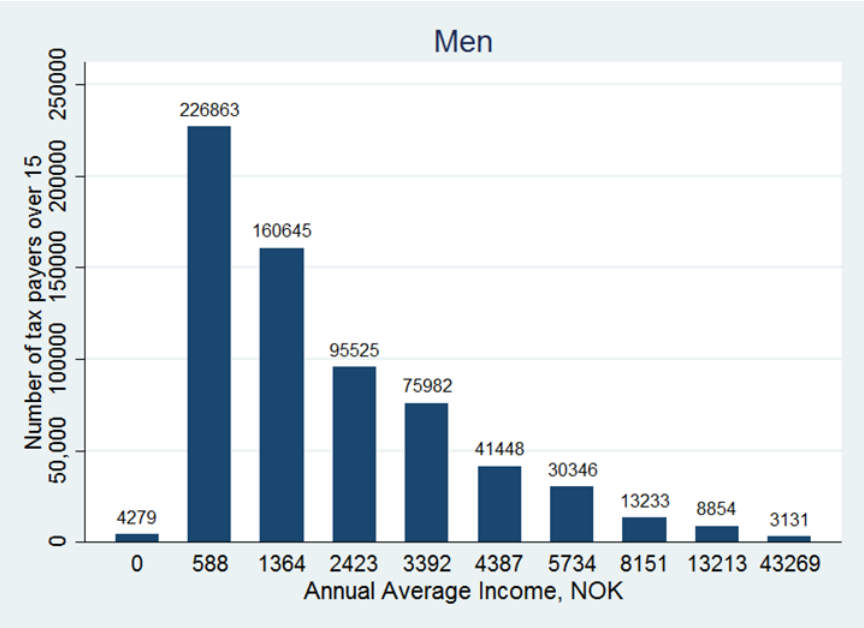
The following figures show income distribution in absolute numbers for all tax payers and various subgroups. The income distribution figures in the main section are in percentages. These figures give us the number of tax payers across all income groups.

Figure A.1: Income distribution for all tax payers



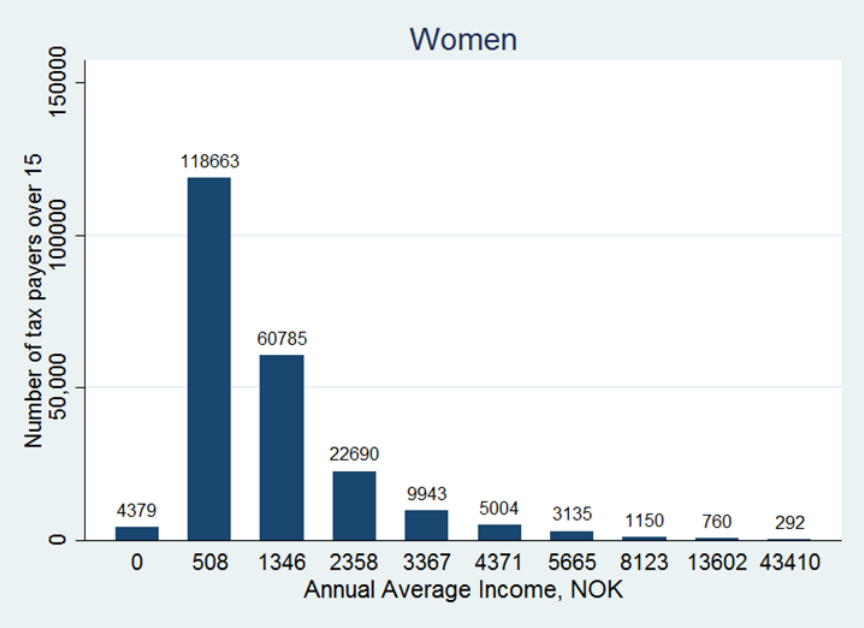
Source: Statistics Norway

Figure A.2: Income distribution for male tax payers



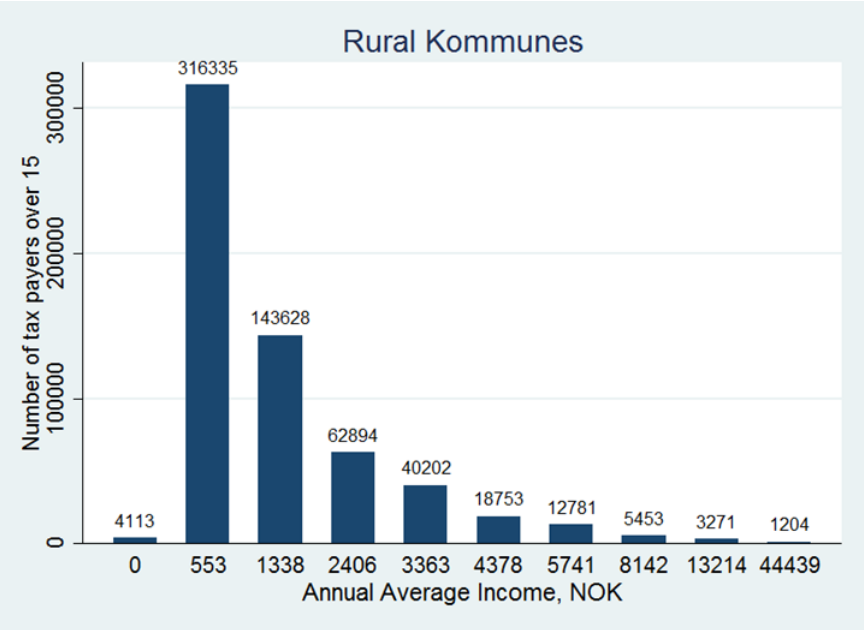
Source: Statistics Norway

Figure A.3: Income distribution for female tax payers



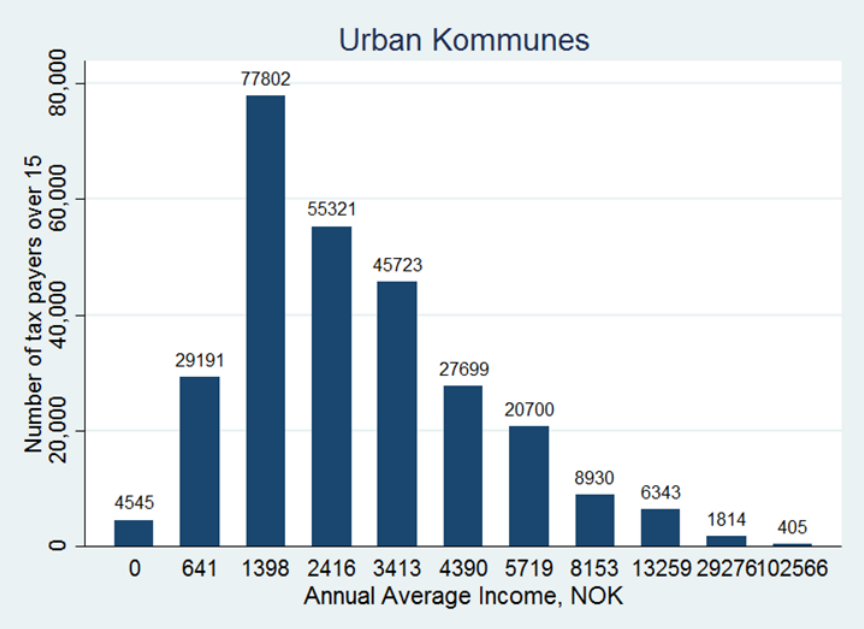
Source: Statistics Norway

Figure A.4: Income distribution for rural tax payers



Source: Statistics Norway

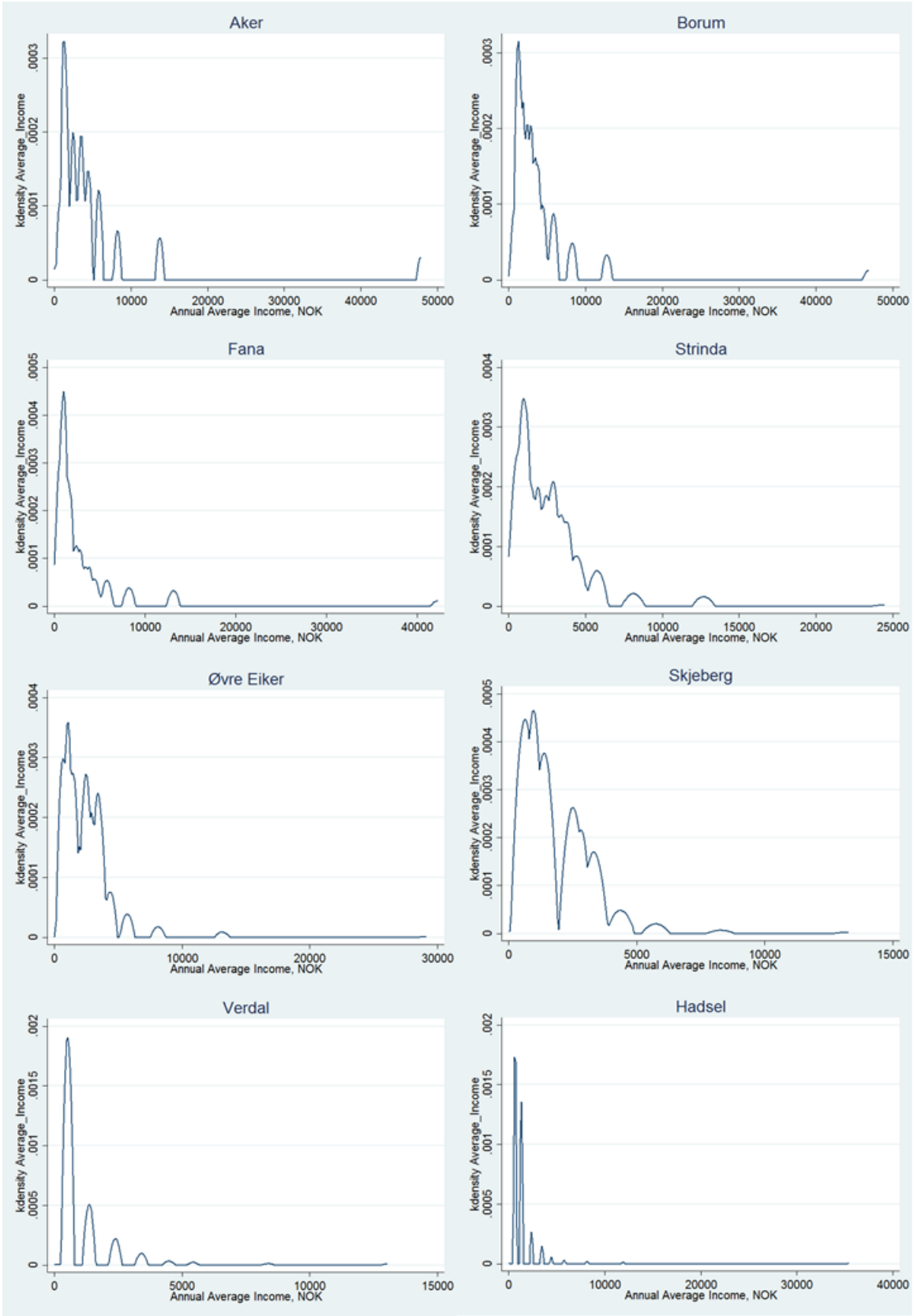
Figure A.5: Income distribution for urban tax payers



Source: Statistics Norway

B Appendix: KDE in Rural Kommunes

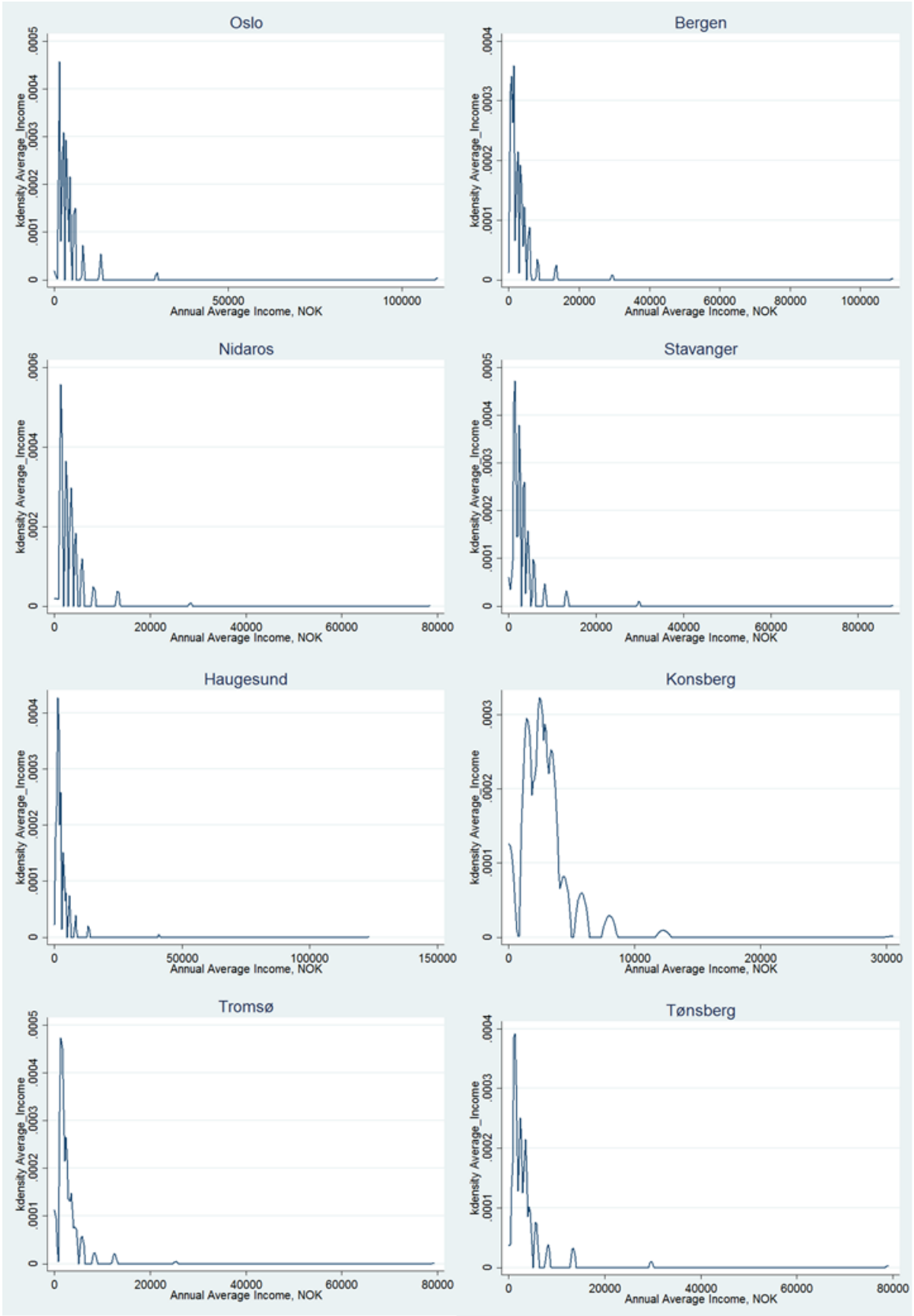
Figure B.1: KDE in 8 major Rural Kommunes



Source: Statistics Norway

C Appendix: KDE in Urban Kommunes

Figure C.1: KDE in 8 major Urban Kommunes



Source: Statistics Norway