Fairness in Nurse Rostering Problem

by

Antonios Glampedakis

The thesis is submitted in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth.

First Supervisor:

Dr. Djamila Ouelhadj

Second Supervisor:

Dr. Dylan Jones

27 September 2016

Declaration

Whilst registered as a candidate for above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate, and has not been submitted for any other academic award.

Dedication

I would like to thank Dr Djamila Ouelhadj my first supervisor, for her help, guidance and patience for the duration of this PhD. My special thanks to Prof. Dylan jones for the valuable insight and wisdom. Also, to Dr Simon Martin for providing me with necessary tools for this research.

To Tzitzi

Abstract

Many Operational Research (OR) problems like scheduling and timetabling, are associated with evaluating the distribution of resources in a set of entities. This set of entities can be defined as a society having some common traits. The evaluation of the distribution is traditionally done with a utilitarian approach, or using some statistical methods. In order to gain a more in depth view of distributions in problem solving new measures and models from the fields of Computer Science, Economics, and Sociology, as well OR are proposed. These models focus on 3 concepts: fairness (minimisation of inequalities), social welfare (combination of fairness and efficiency) and poverty (starvation of resources). A Multiple Criteria Decision Making (MCDM) model, combining utilitarian, fairness and poverty measures is also proposed. These measures and models are applied to the nurse rostering problem from a central decision maker point of view. Nurses are treated as a society, trying to optimise nurse satisfaction. Nurse satisfaction is investigated independently from the hospital management, forming two conflicting criteria. The results from different measures cannot be evaluated using cardinal measures, so MCDM methods and Lorenz Curves are used instead of a numerical, cardinal measure.

Contents

List of tables

List of figures

Chapter 1 Introduction

1.1 Background and motivation

Many operational research (OR) problems are in fact judging the distribution of certain resources. Problems like nurse rostering (eg), air flow control (Bertsimas and Patterson 1998), network flow (eg Bonald and Massoulié 2001, Mazumdar 1991) are in fact evaluating societies in a perspective of either total utility or the fairness of equal distribution. Thus, a social system that can be described as mathematical problem can be subject to fairness, depending on decision makers preferences. Since there is not much research done in evaluating societies in the field of OR, the measures of evaluating societies ought to be taken from other sciences, namely social sciences and economics.

Two different philosophy schools of judging a potential distribution exist: the utilitarian one (Mill 1863, Jeremy Bentham 1789 amongst others) judging an allocation by the total amount of "pleasure" that is distributed in society members, and egalitarian (Lorenz 1905, Arrow 1983, Sen 1973 amongst others) that focuses on the minimising inequalities. Note that most of the egalitarian measures do target absolute fairness, but since it is rarely possible to achieve strong egalitarianism they focus on minimising inequalities.

Societies can be evaluated on frequency distributions of an attribute, which will be referred as utility. Utility can be either negative (for example penalty, constrain violations, time units on hold), or). Every one of those measures can be transformed to utility using a utility function.

In the first section, a number of social choice measures with emphasis on egalitarian ones is going to be listed. The properties that the measures described above must have and calculate the possible computation cost of each one are going to be underlined. A number of widely used measures might be discarded for not fulfilling the necessary properties. A number of them might be discarded due to the lack of computational cost feasibility.

However, judging a society with aspect to only of one of those approaches is, most of the times, incomplete. Even though certain methods can take both utilitarian and egalitarian perspective into consideration (for example, Generalised Lorenz Curve), most of existing

measures focus on either one. Thus, goal programming methods can be used to aggregate those two different approaches.

Merging different approaches from different disciplines has always been a part of OR. In this thesis the aim will be to merge OR with the fields of Social Sciences, and Economics, moving forward in a direction that will help us have a complete view on problem solving.

1.2 Research Objectives

The main aim of this thesis is to investigate the use of fairness concepts in an OR problem. While fairness concepts have been used in OR in the past (e.g. Vasupongayya and Chiang 1995, Muhlenthaler and Wanka 2012) the models used were in essence inequality aversion. In view of this, a distinction between fairness, equity and equality must be made.

Fairness has also be used in MCDM environment. In Romero 2001, 2004 the L_∞ meta-objective, minimising the maximum unwanted deviation in Extended Lexicographic Goal Programming formulation, is a MinMax/Rawls principle that has been a key principle in economics. This opens the question if further inequality averse models can be used in MCDM.

Given those issues, the main objectives of this Thesis are:

- A framework that incorporates concepts of inequality from literature into OR fairness.
- Experimental use of inequality models into a OR problem.
- Explore different approaches in social sciences literature that can be used in OR (e.g. poverty).
- Explore opportunities using Fairness concepts in MCDM.
- Identify ways of evaluating results with regard of fairness.

1.3 Contribution

The main contributions of this thesis are:

- A unified framework for defining fairness. Models are examined on the basis of certain properties, computability being one of them.

- A series of models that are taken from diverse fields of work. Models from economics (Gini Index), computing (Jain index) and social sciences are incorporated in the context of an OR problem.

- In order to evaluate and compare the different results produced by the proposed models a framework for evaluating results with the notion of fairness including graphical representation was used.

- Representing nurses as criteria, ELECTRE method was used to compare models results as well. The weights used in ELECTRE were produced from a formula in order to demonstrate a sense of fairness.

- Introducing models with the concept of "resource starvation" or poverty as objective functions. Those models require the existence of a threshold, poverty line, based on which a member can be "impoverished" or needing more resources

- Measures that incorporate both fairness and equality properties

- Tests that demonstrate the dominance of some models over the others, using the nurse rostering problem. The results were evaluated using the constructed frameworks

- A Goal Programming model using meta goals was introduced. The meta-goals of Fairness, Efficiency and Poverty were implemented. This showed that a formulation combining goals could work.

1.4 Nurse rostering

Healthcare workforce scheduling has been a topic of discussion in academic and non-academic circles the past decades. With the reduction in the number of medical jobs available in the NHS, new issues might arise. Happiness in workspace is also important to productivity (Harter et al 2002). Scheduling and fairness has been pointed to be an important factor (Mueller & McCloskey 1990, Kovner et al 2006, Hayes et al 2010) to nurses' satisfaction.

The Nurse Rostering Problem (NRP) is a complex combinatorial optimisation problem that has been thoroughly studied in the literature. In its general form, NRP is defined as assigning nurses over given shifts over a period of time, subject to a set of constraints. Because of its complexity, NRP is usually approached using heuristic and metaheuristic methods (Burke et al 2008, Martin et al 2013)

Approaches to NRP vary, from modelling the problem (Warner and Prawda 1972, Vanhoucke and Maenhout 2007, Bilgin 2008, Burke et al 2008) developing methods and procedures (For example, Berrada et al 1996, Burke et al 1999, 2001, Constantino et al 2011, Bilgin et al 2012, Martin et al 2013), or incorporating fairness features into the problem (Martin et al 2013, Ouelhadj et al 2013, Wang et al 2014).

The amount of literature (Ernst et al 2004, Burke et al 2004) and the vast attention NRP has received shows the complexity of the problem and the different approaches within the scientific community, as well as the real world application. The complexity of the problem allows for different elements to be considered for every nurse. This may create a lot of diversity in the happiness for each nurse, that in turn allows for diversity in the results produced by the fairness measures. That is particularly important because there are measures ranging from 0 to 1, and a small change could be critical. Since NRP may involve actual humans, examining this problem under the prism of fairness can be a logical development.

In NRP it is common to be modelled as constraint violation problem. There are several types of constraint violations, we can group them in 2 different categories: Hospital needs, and nurses' needs/preferences. The fairness concept on this work focuses on nurses: instead of adding all constraint violations for the group of nurses, constraint violations for each nurse is calculated. This way each nurse has a negative "utility" upon which fairness concepts can be implemented.

Further information for the benchmarks and the NRP model used will be presented in Chapter 4.

1.5 Outline of thesis

The thesis is structured as follows:

Definitions and concepts of fairness are presented in chapter 2. A review on the literature on philosopher work about equity, fairness and equality is provided. Connections with applied sciences like economics or social sciences are made.

In chapter 3 the models are presented and examined. Fairness, efficiency, poverty and Social welfare function (SWF) models are explained and examined based on their properties. An algorithmic analysis is also provided. Fairness in operation research problems is also investigated.

In chapter 4 some tools that are used throughout the thesis are presented. Nurse rostering problem is introduced, as well as the benchmarks used. Lorenz curves and ELECTRE are presented. Goal programming is also described.

In chapter 5 poverty concepts are introduced. A framework is established for poverty measurement and identification. Poverty measures are evaluated based on the aforementioned framework.

In chapter 6 the adaptation from fairness measures for maximisation problems to minimisation are made. Results on fairness measures are presented. Evaluation is performed using ELECTRE and Lorenz curves.

Chapter 7 presents results for poverty measures. An analysis on different poverty line is being made. Results are presented and visualised with a number of different techniques.

In chapter 8 a combination of fairness and efficiency measures are introduced, as social welfare functions, with a goal to optimise in both the quality of roster and minimisation of inequalities. Again, nurses and hospital needs are examined.

In chapter 9 Goal programming model is introduced (Jones and Tamiz 2010). The parameters that are used affect both the nurse roster (efficiency, fairness, minimisation of unhappiness) and ward coverage. Results show the use of having multiple objectives rather than just one.

The final chapter reviews the achievements of this thesis, presents general conclusions and propose directions for future research on fairness models methods and applications.

Chapter 2 Fairness introduction

2.1 Introduction on fairness

The topics of fairness and equality are concepts that engage a diverse set of disciplines. Philosophers, politicians, religions, economists, social scientists, and recently scientists that are connected with Operational Research (OR) (e.g. computer scientists) have been occupied with it. In this chapter the general concept of fairness, equality and equity are introduced. Differences between fairness, equality and equity are presented and point out the diverse set of disciplines that fairness is connected.

2.1.1 Fairness

Oxford dictionary defines fairness as "Impartial and just treatment or behaviour without favouritism or discrimination."

The concept of fairness is met since the earliest philosophy literature, beginning in Ancient Greece. In ancient Greek the word «δίκαιο» meant both fair (in the ethical way) and just (in a lawful way), and the concept of it was encountered in a wide range of things, from the government of communities to transactions.

Plato: "the republic" –πολιτεία- it is underlined by Polemarchus that ascribing equals is fair. A cynical view that was described was that the stronger or richer people should be treated fairly, or better than others, since they can afford it. Plato describes how a community would prosper, by introducing a concept of decision makers, the philosopher-kings. The state should be run by those experts that their goal is the good of the whole community and balancing the needs of the several members of society with the good of the community in mind. In this description the state is fair in scope of a community, even though some individuals may not be as privileged.

Marx (1875) outlined two different cases of fairness: need based fairness and work based fairness. In the stage described as socialism, everyone is rewarded depending on his contribution to society. In the stage described as communism, people are being treated based on their needs. This is because in communism resources are abundant there are no restrains on the amount of resources distributed.

2.1.2 Equity

Equity concept is common in philosophical discussions. Oxford dictionary defines equity as "The quality of being fair and impartial".

Hobbes (1588-1679) presented a definition of the equity concept. He stated that equity is a human quality that characterizes everyone or some individuals as well as it is a law of nature that humans are obliged to follow.

According to Kant, one can't demand being treated favorably. However, if someone claims something on the basis of equity, he is obliged to treat others like that. Kant gives a number of examples of equity, in different contexts, that recite on the concept of equal pay for equal work. For instance, in a company where the shareholders have different participation in capital, the profits or losses that incur on individuals should be proportionate to their participation.

Early Philosophers like Bentham (1789) discussed about the utilitarian approach. Bentham advocated that the measure a societies total happiness is the sum of the happiness of individuals. Any increase on total happiness would lead to a better society. This approach has been criticized since it takes in no consideration the distribution among individuals.

2.1.3 Equality

Oxford dictionary defines equality as "The state of being equal, especially in status, rights, or opportunities."

According to Dalton (1920) states that equality depends on the context that the definition is applied. An example is presented in an economic context, where the situation of perfect equality is reached when the total income is distributed in equal parts among a number of persons.

As it is presented above, the concept of equality is often treated in regard to the equity concept. In the following paragraph this aspect is focused pointing out the differences between equality and equity.

According to Bronfenbrenner (1973) equity is a subjective concept whereas equality is objective. He describes the differences between equality and equity and concludes that even though there are phonetic similarities and philological connections, the two terms are quite

distinct. He mentions that the equity is non –mechanical in principle and is in essence a subjective matter. In order to achieve equity, the wealth distribution has to be done in line with principles of justice.

Equality is largely a mechanical matter and in fact is associated with a measure that can be equal, such as wealth per unit or income. In addition, Espinoza (2007) states that while equality involves a quantitative assessment, equity has an ethical judgment and a quantitative assessment as well.

Equality can also be described as a state that no individual wants to take the place of another. This is also the concept of Envy-freeness, where comparisons between individual utility is not relevant, and an allocation is considered envy free if no one prefers the state of another individual.

Rousseau (1954) claimed that inequality is inherent to societies, and related to the concept of property. However, men were created as good, and with no ill will but society corrupted them.

It is commonly presumed that fairness, equity and equality are identical. However, even when a system is fair it can create inequalities, depending on constrains and individual utility functions. Measurement of inequality is usually done by a series of statistical dispersion measures.

2.2 Fairness in economics and social sciences

Sen (1982) introduce the questions of "equality of what". Sen attempts to give an answer to this question by stating that we have to be preoccupied with the distribution of capabilities in order to reach valued functioning. Instead, Rawls (1971) states that we should be concerned with the distribution of primary social goods. Even though Sen is interested in equality of capabilities this doesn't confute Rawls answers about primary social goods because how wealth, income, power, status, education, work is distributed has an effect on people's capabilities. Other approach has taken by Arneson (1989) who mentions opportunities for welfare, Dworkin cares about allocation of resources and also Cohen (1989) favors access to advantage.

Dasgupta (1993) presented equity among individuals or groups as a measure of the relative similarity when the groups or individuals enjoy material resources, education, socio –political

rights, health, education and technologies. It is concluded that equity is accomplished when each group gets its fair share.

Rawls (1971) outlined the basis of a fair system: Everyone is entitled to basic liberties, is considered equal to others for that matter, and the greatest benefit of the least advantaged members of society is most important for fairness-pursuing policies. Rawls proposed a fair procedure that it would lead to a fair distribution of primary and other goods. His procedure based on the hypothetical scenario that a group of persons have to reach an agreement about their political and economic preference for the society. The proposed procedure argued that the final allocation of primary goods and recourses give an egalitarian distribution of outcomes.

Instead of evaluating the distributions according to actual income economists (Dalton 1920, Sen 1973 1980) proposed calculating the utility that income will bring to each individual. Then evaluate the solution based on utility.

Apart from fairness, efficiency is the most common way to describe a solution. It is usually measured by a central tendency measure, usually the average. Thus, distributions are not only evaluated about how spread is the personal income, but how much is to be distributed. This is the concept of Social Welfare, a combination of efficiency and minimisation of inequality, a central idea in economics. In this essence, we can't be fair in a society level if we accept this society to administer less utility than we could. However, Sen (1970b) proved that having maximisation of goods under total fairness is not always possible.

Other theoretical work includes Atkinson (1970), where he induced a set of inequality-averse functions, the a-fairness. This class of measures ranged from completely inequality-aversion to utilitarianism.

Lambert (1993) using Social welfare utility function discussed tax policy and tax reforms. In his final chapters, he discussed the impact on of tax to the entirety of the population. The distributive consequences of needs-based approaches in redistributing utility are also discussed.

Nash equilibrium (1950a, 1950b) has a notion of fairness between competitive individuals, where an allocation is improved if the percentage improvement of one individual is greater than the percentage decrease of the other individual.

Apart from numerical measures, outranking methods such as Lorenz curves or voting methods can be used for determining fair solutions. These methods cannot guarantee a single outcome as index measures do, but require a minimum level of assumptions. Lorenz and generalised Lorenz curves will be reviewed in a following chapter.

2.3 Connection with OR

In order to justify the use of fairness concepts in OR problems, an analogy between them must be made.

Society has a number of different definitions. Oxford dictionary defines society as "The aggregate of people living together in a more or less ordered community." WorrdWeb defines society as "A formal association of people with similar interests" and "An extended social group having a distinctive cultural and economic organization"

In that sense, every OR problem that includes a set of different entities can be a "societal" problem, and so be subject to fairness or other society related approaches.

Entities in OR problems that could take the part of individuals, and thus making an OR problem societal may differ immensely. Those entities could be actual human beings constituting a society like nurse scheduling (Warner 1976). In facility location problem (McAllister 1976) where the subject of fairness could be groups of individuals that require access to a public facility. In portfolio analysis (Iancu 2014) where the investors split the market impact costs in a fair way. Even in job shop scheduling (Sabin 2004) fairness has been applied in accordance to scheduled jobs. The aforementioned research states that jobs that might take longer to finish are being "discriminated" compared to shorter ones and deals jobs in a fair way. There are different kind of individual or entities that could be considered in aspect of fairness in different OR problems.

For example, in nurse rostering every nurse is part of a hospital, or a hospital ward. Even though nurses operate with different contracts and might have different preferences they work in the same space being restricted by similar rules and working conditions. Nurses have the more or less the same preferences on working conditions which they may vary, however they are more or less standard (Kovner et al 2006, Hayes et al 2010, Coomber 2007).

A most important issue is the value upon which societies are evaluated. In Economics, a currency is used to determine fair or efficient distributions. For example, in nurse rostering

constraint violations is a negative "Utility" to be minimised. In job scheduling (Vasupongayya 2005) waiting time (among others) is a negative "Utility" to be minimised. In facility location, utilisation of a facility is among others, a "currUtilityency" to be maximised.

Social sciences examine the distribution of a quantity in a population comprised as a society. This quantity can take the form of happiness, access to healthcare, employment, quality of life etcetera. In economics, this quantity is usually a form of currency for example, Atkinson (1970).

In both cases, this quantity represents a utility $U(x)$, where x is the individual, and the abundance of utility is preferred to scarcity ($maxU(x)$). The purpose is to evaluate societies in regard to the distribution of this utility.

For example, in p-median facility location problem, the various entities are assigned to a hub. Members comprise the society. Distance between member and the hub is the quantity to be optimised, thus the utility. The general direction is to minimise the distances between the hubs and the members assigned to them, ($minU(x)$).

A similar comparison can be made in a series of OR problems.

Some OR problems are minimization problems. Utility in those cases is considered to be a negative quality ($minU(x)$). This simply means that the individuals with lower negative utility are better off.

Arrow (1950) proved that under a specific set of rules a completely fair system is not always possible to exist, and that may result to a "dictator" or a central decision maker. The will of the central decision maker can be simulated by an objective function. This objective function should include all potential goals of the decision maker, which can include fairness goals.

In OR, fairness is confused with inequality. Even though the premise of equity is present, when fairness is mentioned, the goal is most of the times minimisation of inequalities (eg, Ouelhadj et al 2012)

2.4 Conclusion

In this chapter the basic notions of fairness, equity and equality were presented. Some examples of those concepts were presented in social science and economics disciplines. Differences in interpretations on every term were demonstrated. From the number of

interpretations in every case, it can be inferred that those concepts depend on the context. A groundwork of applying fairness concepts to OR problems was laid.

Chapter 3 Fairness in Operational research

3.1 Introduction

In operational research (OR) fairness is mostly interpreted as minimisation of inequalities. The research is mostly towards a set of OR problems such as scheduling, allocation, and facility location problems.

Societal problems may include a large array of OR problems. Scheduling and allocation problems are the core of this set. The analogy between society and societal problems can be made on the basis that an element is distributed in OR problems the same way it is distributed in society. In society, economists and social scientists study about utility and welfare distributions. Usually utility is positive in the form of currency or some other beneficial value such as access to clean water. There are cases that utility is negative, such as child mortality (Rutstein et al 2004). In OR the positive utility that is to be distributed can take form of bandwidth in network allocation, completion in work scheduling, or even income in portfolio analysis. Some negative utility distribution cases are constraint violations in crew scheduling problems, or traveling tournament problem.

Based on those principles, and on the premise that every set of individuals with a common characteristic or a common reference point can represent a society, it can be can evaluated an allocation based on either egalitarian or utilitarian characteristics.

Until the recent years, the norm of evaluating allocations is based on Utilitarian approach. The most common concept was that the preferred allocation was the one that maximised the resources allocated. Some examples on this approach on societal problems are Workforce scheduling (Warner and Prawda 1972) location problem (Church and Revelle 1974) airflow management (Bertsimas and Patterson 1998). In those cases maximising the average, or just the sum of allocated resources was the objective function of the problem.

In OR the definition of fairness is vague. It usually takes the form if preventing inequalities. For example, in Rafaeli et al (2003), fairness in queues is considered the "perception of fairness" in individuals as the feeling each individual gets for waiting in queue. In Ee (2004) fairness is stated to be as resources in each node are "approximately equal".

The study of how the resources are distributed was effectively the egalitarian approach. In certain problems, the literature is bigger than others. In facility location problem a review of

the existing literature in fairness-inequality measures can be found in Marsh and Schilling (1994).

Rawls criterion (1971), for others max-min criterion or Chebyshev max-min, is the commonest fairness measure. It was first expressed in economics, and it is common in a number of applications in OR such as Nurse Rostering (Ouelhadj et al 2013) Academic timetabling (Mühlenthaler and Wanka 2013), Network allocation (Bonald and Massoulié 2001, Arulambalam et al 1996), Congestion control (La Boudec 2005) and water filling (and Le Boudec 2006). Max min fairness concept is to improve the worst-off member of the society. This has certain limitations. Judging a distribution by a single individual disregards the majority of members in a society, and the shape of the distribution. Also, expressing a result by a number that is not connected to a form of centre measure, disregards efficiency in the final outcome. It does not necessarily link to Pareto optimal solutions, and it is not scale invariant (see section 2.1).

Another way of formulating equality is to consider equality as a form of deviation from some point, usually the average. For that cause, a series of statistical measures can be used. Standard deviation is the simplest of those. However, it does not comply with strong transfer principle. Variance on the other hand, is satisfying Pigou-Dalton principle (1920). It has been included in a multitude of OR problems as a fairness measure such as income distribution (Atkinson 1970) Facility location (Berman et al 1990, Mcallister 1976) Job scheduling (Vasupongayya and Chiang 2005), Network allocation (Jain et al 1984). However, it is not scale invariant, and not bounded (see section 2.1).

Apart from every statistical measure such as Deviations, Variance, Range (Gopalan and Batta 1980, Muhlenthaler and Wanka 2012), Max-min, normalised versions of those measures can be found, usually divided by average. This creates a series of bounded measures (Marsh and Schilling 1994 has a categorisation of measures) that is a very useful property in this form of Social welfare measures (see section 2.1)

Gini coefficient is a popular measure in economics (Gini 1912), with an applied effect in income and tax analysis, but also in social sciences (Daly et al 2001) that is based on Lorenz curves. Applications in OR are diverse, Facility location problem (Drezner et al 2009, Yapicioglu and Smith 2011), healthcare applications (De Bruin et al 2010), and portfolio optimisation (Ringuest et al 2004).

Jain (Jain et al 1984) index is a popular measure in OR. It was introduced in network allocation, and has applied to personnel rostering (Ouelhadj et al 2012), academic timetabling (Muhlenthaler and Wanka 2012), Network allocation (Drougas and Kalogeraki 2005, Shin and Lee 2005, Tang 2010 Belleschi et al 2011 etc). It is also used as a comparative quality measure between solutions produced by other fairness models. It is popular mainly in computer science related disciplines.

Atkinson (1970) set of measures (or A-fairness) is a set of utility social welfare measures. The difference from previous statistical measures is that incorporates efficiency as well as inequality measurement. Depending on how you formulate it, it includes max min fairness, average and Theil's measure. It provides the decision maker with a choice of how much inequality averse or efficiency averse he wants to be. As social welfare measures, they have been used in cases that a total evaluation of the solution is needed Air flow management (Bertsimas 2011a), Healthcare applications (Bertsimas et al 2013 Leach 2011), Facility location (Yang 2011)

An analysis for every measure is provided in Section 3.4

3.2 Proposed Framework

The majority of research work has been done to clarify the ways of evaluating societies. The differences between inequality, efficiency and social welfare have been studied. A list of measures has been investigated according to a set of properties.

There have been many efforts to set a framework of properties for fairness measures (Bourguignon 1979, Trichakis 2011, Joe-Wong et al 2011). Lee and Chiang (2010) proved that the only measures that comply with a certain number of properties including strong transfer principle were either logarithmic or power ones.

In order to evaluate measures that are to be used in evaluating societies, a set of properties is introduced to form a framework. Depending on the problem, a fitting measure could be either social welfare measure or fairness one. In case of OR most problems need Social Welfare instead of fairness measures, in order to incorporate efficiency. Thus, the framework that is proposed consists of the following properties:[1]

---

[1] In cases of fairness a needed property would be scale invariance and a desired one would be boundedness. Scale invariance ensures that the choice is invariant if every individual's income is

Needed properties

- Transfer principle (Pigou-Dalton Principle)
- Anonymity (Symmetry)
- Boundedness

Desired properties that are used to evaluate measures include:

- Decomposability
- Computational feasibility
- Principle of population
- Scale invariance

The transfer principle (Dalton 1920, Pigou 1912) indicates that every transfer that improves a poorer person in the expense of a richer person leads to less inequality. In the same way, every transfer from a poorer person to a richer one makes society less equal. So, in two different ordered distributions, A, B, where $A = (x_1, \dots, x_l, \dots, x_i, \dots, x_n)$ and $B = (x_1, \dots, x_{l'} + l, \dots, x_{i'} - l, \dots, x_n)$, where $0 < l < x_i - x_j$ is a positive value, and x_k are income values for the n people A would be considered less Equal than B.

Anonymity (symmetry) ensures no inherit discrimination by any property. It implies that the identities attached on members of a distribution pay no role whatsoever in distribution comparisons. This is necessary even in cases that some people deliberately undermine their position in favour of the others. In such distributions, each person's view about Utility or penalty is calculated before the final evaluation of the solution. While anonymity is a property mentioned in all the aforementioned literature (eg Bourguignon 1979, Joe-Wong et al 2011) all equality measures seem to satisfy it. Therefore, anonymity is the property that ensures the existence of Fairness-equity: Everyone should be treated equally, regardless of status.

While it might be a controversial property, in the sense that some people might argue that some historical or other reasons are justification to treat people differently. This is more the subject of what fairness investigates. A small discussion is mentioned in 2nd chapter. In a

---

multiplied by the same amount. Boundedness is when a fairness measure values have an upper and a lower bound. In the process we evaluate fairness measures according to these propertie as well.

context of a specific problem, some amendments can be made in order to incorporate personal views. For example, in nurse rostering senior nurses, or nurses with families and therefore special needs, are treated differently than entry nurses, or nurses with no obligations. While this is always a matter that a decision maker should examine. A common solution is that depending on nurse's needs, more or less constraints are imposed, and accordingly the value of a constraint violation might be different.

Computational feasibility is referred to a measures computational cost: the ability to calculate the measure in a reasonable amount of time. In most approaches in OR and in this thesis specifically, a search function is performed a number of times calculating the objective function. In some cases (also see Chapter 6) calculating the objective function might be a large part of the time surpassing the time needed to perform the search function. In the scope of this work, each measure is run a specific number of times for two reasons: so that the effect of the measures complexity can be compared with the total run time and so the computational complexity of each measure does not account as a detriment to the quality produced by that measure. Insight for the calculation of computational feasibility is given in 3.4.

Decomposability states the worth of a distribution can be figured by calculating the worth of parts of the distribution. Additive decomposability will occur with the sum of mutually exclusive parts. Given 2 distributions $A(a_1, a_2, ..., a_i, a_{i+1}, ... a_n)$ and $A'(a_1, a_2, ... a_i, a_{i+1}+k, ... a_n)$ with a subset of distribution same to both allocations A and A', and the remaining members have less inequality in A than A', then distribution A is more equal than distribution A'. In cases of heuristic search methods, where the distributions occur with a small difference from a previous distribution, calculating the measure for a part of the distribution and then compose it with the rest may prove faster than calculating the whole measure from the start.

The principle of population states that in case we scale the population in a distribution, creating a new one with more members of the same Utility then equality index would remain the same.

By boundedness, Jain (1984) states that a fairness measure must be between 0 and 1. However, any measure with upper and lower limits can be normalised to [0,1] This way fairness can be expressed as a percentage. Fairness of 1 could mean a total equality, while values near 0 imply total inequality. This way, conclusions about the fairness of a distribution

can be made in a more intuitive way. If a fairness measure is bounded, a combination with an efficiency measure may lead to a more complete understanding of the distribution.

With this framework, a few remarks can be made. A fairness measure is a Lorenz-consistent, bounded, continuous, and additively decomposable inequality measure if and only if it is a positive multiple of a generalized entropy measure (Shorrocks 1980). Bourguignon (1979) proved a series of theorems for decomposable and symmetric measures. Lan et al (2010) proved that the only measures that comply with a set of properties (weak transfer principle, continuity, scale and population invariant and decomposable) are the ones with logarithms or power functions.

3.3 Measure analysis

Statistical measures

Several statistical measures have been proposed in the literature. Several of them are easy to compute, however do not satisfy some of the properties proposed in the framework. In all cases below, y_i denotes the currency or utility. In nurse rostering context y_i denotes the total weight of constrain violations for a specific nurse. n is the size of the population while μ is the average utility in population. In nurse rostering it is the number of nurses and the average constrain violation of the nurses' population respectively.

Average

$$avg = \mu = \frac{\sum_1^n y_i}{n}$$

The average has been used as a utilitarian measure (countries GTP), as well as a complimentary measure to an egalitarian one. In cases that the number of members that the Utility is distributed is unchangeable, the average measure can be reduced to just a sum of Utilities. In cases that the number of people in the population changes, average can be used as an easy way to measure total Utility, with accordance to the population. The distribution with the highest average always is Pareto optimum.

Used as an egalitarian measure, the average is neither scale nor translation invariant or satisfies the transfer principle. Its use as a central-efficiency measure is dominant, but there are arguments for other central measures, such as median or mode, as well.

Max-min (or min-max)

$$W = min(y_i)$$

Rawls (1971) proposed this measure, with the goal to always improve the worst member of society. This measure targets the worst-off member of the distribution, in a sense that "A dollar is a dollar for the poorest".

A transfer from a richer member to anyone but the worst-off member does not contribute to equality. On its own it also allows wastage of resources. Max-min is not scale invariant, nor bounded: If the utilities of the population is multiplied by a fixed number, then the evaluation of society according to Rawl's measure will change.

In cases that total Utility is a fixed number, this measure can sometimes lead to total equality. However, in cases of large problems using heuristic methods this measure may not optimise towards the right direction.

Standard deviation:

The Standard deviation of a data set is the absolute difference between that element and a given point. Typically the point from which the deviation is measured is a measure of central tendency, most often the mean (average) or the median point of the data set.

$$D = \sqrt{\frac{1}{n}\sum_{i=1}^{i=n} |y_i - \mu|}$$

In either one of those points of measuring deviation, the transfer principle is not strongly satisfied. If utility from a richer person is transferred to a poorer person above the mean the absolute deviation remains the same. This measure is translation invariant, but not scale invariant. In comparison to variance, its sensitivity to large transfers from richer members to poorer members is decreased.

Variance

$$Var = \sigma^2 = \frac{\sum(y_i - \mu)^2}{N}$$

This is a statistical measure that has been used as a fairness measure. It is not scale invariant, since if every utility is multiplied with a fixed number the differences between them will

increase. It does satisfy other properties, like the transfer principle and the population principle. Values near zero signify larger inequalities in the distribution, while values near one smaller.

This measure has the following nature: the transfer of Utility from a "richer" member to another member that is above average leads to a smaller difference from the transfer from the richer person to someone that has lower Utility than the average. This is a desired property in equality measures.

Range (Difference between best and worst individual)

$$R = max y_i - min y_i$$

Range is one of the simplest measures of distribution dispersion. It is the difference between the worst-off and the better-off member of society. It does not satisfy the Pigou-Dalton principle. This measure is translation invariant, but not scale invariant.

Even though it is easy to calculate, this measure only takes into consideration the two most extreme members of a society and it can be argued that it is inefficient.

Gini coefficient - Generalised gini coefficient

$$G = \frac{2 \sum_{i=1}^{N} i * y_i}{n \sum_{i=1}^{N} y_i} - \frac{n+1}{n}$$

Gini's index is a measure directly derived from Lorenz curves (see 4.3): it is the area between the Lorenz curve of the distribution, and the equality line. It ranges between 0 and 1, with 0 being total equality. Since it is based on a sorted allocation Gini index satisfy transfer principle, scale invariance and population invariance. However, it is not generally additively decomposable and sorting methods are expensive computationally. An example of graphical representation of Gini index would be in the following graph 3.1.
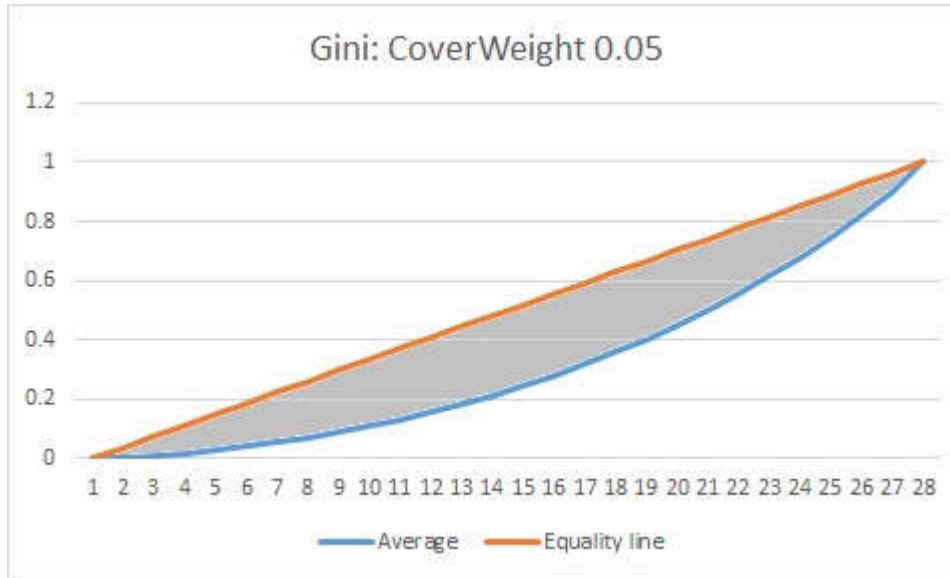
Figure 3-1Gini index

Gini is one of the oldest proposed inequality averse measures in literature proposed since 1905. There is a lot of research about calculation of Gini index (eg, Pyatt et al. 1980, Milanovic, 1997, Gastwirth 1972). The above calculation derives from

$$G = \frac{1}{N}(N + 1 - 2 * \left(\frac{\sum_{i=1}^{N}(N + 1 - i) * y_i}{\sum_{i=1}^{N} y_i}\right))$$

and it is simpler in calculation terms. N again stands for the population, y_i is the income of the sorted individual i.

Another way to view Gini index is as a function of a covariance

$$G = 2 * \frac{covar(y, i)}{N * avg}$$

Where $covar(y, i)$ is the covariance between income and ranks of all individuals (Pryatt et al 1980).

For the analysis of Gini index there is again a lot of work in literature. Bourguignon(1979) discussed about the additive decomposability, while Lerman and Yitzaki (1985) discussed about the decomposability of Gini using the covariance formula.

Since Gini calculation through the analytic formula requires the use of a sorting algorithm calculating Gini can be time consuming for larger populations. Work that tries to bypass this issue includes, but not limits to Hoeffding (1948), Glasser (1962), Sendler (1979), Beach and Davidson (1983), Gastwirth and Gail (1985), Schechtman and Yitzhaki (1987).

Jain index

$$J = \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N * \sum_{i=1}^{N} (y_i)^2}$$

Jains index is a bounded measure that is scale and population invariant but not decomposable. It ranges from 1 to $\frac{1}{n}$. One is the most equal solution that everyone has the same resources allocated. $\frac{1}{n}$ would be the most unequal solution and would be close to 0 (how close to 0 depends on the population size) where one individual has all resources. Regarding decomposability, if a subset ui of a distribution U is known and it's Jain index value can be calculated, while Jain's index of a complimentary subset uj ≠ ui is calculated, the Jain's index of the whole distribution can be found as a function of uj and ui.

Regarding other desired properties Jain index satisfies, Jain index is easier to compute than GINI since there is no sorting or some other complex algorithm. It is also scale invariant and satisfies the transfer principle as mentioned by Jain(1984).

Theil's measure

$$Theil = \frac{1}{N} \sum_{1}^{N} \frac{y_i}{\mu} \ln \frac{y_i}{\mu}$$

A not bounded measure that can be transformed to a bounded measure if divided by ln(n). It satisfies every fairness property in the framework, however due to logarithms it has some problems. In case that there are some individuals or groups with zero utility the logarithm could not be computed. Also, logarithms are expensive computationally wise, and might prove ineffective.

The logarithm, however, provide a very interesting property. The changes in the individuals in the lower part of the distribution are more important than the changes in the upper part of the distribution. This allows for a more inequality averse measure, that is targeted in the poorer population.

Atkinson index

$$A_\varepsilon = 1 - \frac{1}{\mu} \left(\frac{1}{N} \sum_{i=1}^{N} y_i^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}} \quad for \; \varepsilon \neq 1, \varepsilon \geq 0$$

$$A_{\varepsilon} = 1 - \frac{1}{\mu}\left(\prod_{i=1}^{N} y_i\right)^{\frac{1}{N}} \text{ for } \varepsilon = 1$$

Atkinson index is a set of fairness measures. Depending on ε it can take the form of a measure that is strongly utilitarian, strongly egalitarian, or anything in between. Some notable cases include $\varepsilon = 0$ where it expresses the utilitarian approach of a measure, or average: every "point" of utility distributed is equally good for everyone regardless how "rich" or "poor they are. For $\varepsilon \rightarrow \infty$ it is a normalised max-min measure, also analysed above.

Since Atkinson indices is a class of inequality measures, depending on the value of ε Atkinson index satisfies or not a number of properties. Decomposability is definite as well as scale and population invariance. Transfer principle depends on the choice of ε. It does not apply on either the extreme values of ε, but it may apply for intermediate values.

The total welfare of a distribution can be represented as a function of welfare of each member (1)

$$W = \frac{1}{N}\sum U(Y_i)$$

Where W is welfare, and the function of $U(Y_i)$ according to Atkinson (1970) could be (2)

$$U(Y_i) = \frac{1}{1-\varepsilon}y_i^{1-\varepsilon} \text{ for } \varepsilon \neq 1$$
$$U(Y_i) = log y_i \text{ for } \varepsilon = 1$$

The base of Atkinson index is the concept of Equally Distributed Equivalent (EDE). EDE is the utility that if every individual had, then the whole population-society level of welfare would be the same as actual incomes total welfare (Atkinson 1970). An example of EDE for a distribution of population of 2 presents to Figure 3.2.
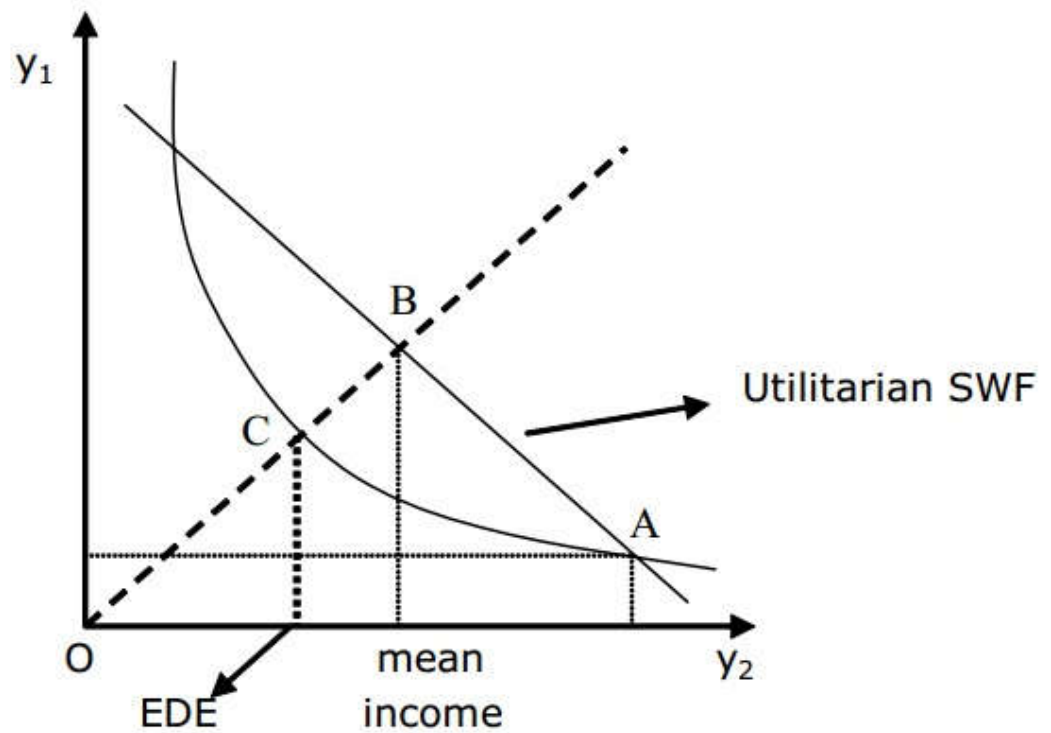
Figure 3-2Graphical representation of EDE

Atkinson index derives from EDE as following (3):

$$A_\varepsilon = 1 - \frac{OC}{PB} = 1 - \frac{y_{EDE}}{\mu}$$

According to (2), $U(y_{EDE}) = \frac{1}{1-\varepsilon} * (y_{EDE})^{1-\varepsilon}$

From (1) and (2) we can get that (4)

$$y_{ede} = \left\{ \frac{1}{N} \sum_{1}^{N} y_i^{1-\varepsilon} \right\}^{1-\varepsilon}$$

From 3, 4 we reach to the Atkinson index of inequality.

3.4 Algorithmic analysis of measures

In order to determine the computational cost and the time needed to calculate a measure, some algorithmic analysis must be made. Algorithmic analysis determines the number of operations in a given algorithm. The difference with big O notation is that it does not only determine an asymptotic approximation, but also the number of total operations for an algorithm. Another reason for doing an algorithmic analysis is that operation costs are

different depending on the programming language used. A computational cost analysis, based on some assumptions has been made. This analysis is provided in table 1, table 2 and table 3.

It is assumed that any of the following processes are exactly one operation (Katajainen and Träff):

- Loading or saving to memory
- Increments by 1
- Additions
- Comparisons

Multiplication and division is $k * n$ operations, where n is the number of digits of the numbers. For the sake of comparisons, it is stated that $k = 3$.

In some of the measures a sorting algorithm must be used. The choice of the sorting algorithm is mergesort – since it is the one used by Java. The sorting algorithm is only used in Gini index and its complexity is calculated to $8\mathrm{N}\log_2 n - 4 * 2^{\log_2 n} + 2n$ (Katajainen and Träff) .

It is assumed that the logarithm is up to the 3rd digit. That makes out 12n operations (Hart 1978).

All computational costs were calculated using the algorithms in our experiments.

Table one is describing simple statistical measures. Table two includes measures from economics and network allocation.

Statistical	Max min	Avg	Max Diff	Absolute Deviation	Variance	Coefficient of variation
Transfer principle	✓	✗	✗	✗	✓	✓
Scale invariance	✗	✗	✗	✗	✗	✓
Principle of	✓	✓	✓	✗	✗	✓
Decomposability	✗	✓	✗	✓	✓	✗
Pareto	✗	✓	✗	✗	✗	✗
Boundedness	✗	✗	✗	✗	✗	✓
Complete ordering	✗	✓	✓	✓	✓	✓
Landau notation	O(n)	O(n)	O(n)	O(n)	O(n)	O(n)
Alg analysis	5n	5n	6n	10n	10n	10n

Table 3-1: Statistical measures

Advanced	Theil	Gini coefficient	Jain Index	Atkinson index
Transfer principle	✓	✗	✓	✓
Scale invariance	✓	✓	✓	✓
Principle of population	✓	✓	✓	✓
Decomposability	✓	✗	✓	✗
Pareto	✗	✗	✗	✗
Boundedness	✓	✓	✓	✓
Complete ordering	✓	✓	✓	✓
Landau notation	O(n)	O(nlogn)	O(n)	O(n)
Alg analysis##	20n	$12n + 8\text{n}\log_2 n - 4 * 2^{\log_2 n}$	10n	10n-3n*e

Table 3-2: Advanced measures

To conclude, measures like Atkinson index are descriptive enough without being costly. Measures that need a sorting function are maybe too costly for implementing in large problems. Most of statistical and advanced measures are not Pareto efficient without using a social welfare function. Statistical measures like average, min-max, or standard deviation cannot be much faster than some more advanced ways of calculating social welfare, like social welfare with Jain's index. However, for more accurate results, an application to problems is needed.

3.4 Search method

The basic simulated annealing meta-heuristic used by the agents is implemented with a choice of a cooling schedule. After finding a feasible starting solution the algorithm moves to possible neighboring solutions. The move is always accepted if it leads to a better roster, and it is accepted with exp Δ/t probability if it leads to a worse solution. Δ is the difference in the objective function, where t is the temperature that decreases with a geometric cooling schedule, set at 0.9, gradually permitting fewer chances in accepting worse solutions. A fixed number of moves are allowed per run. The tests were performed in an HP EliteBook with 8GB ram, 4x Intel i7-3540M processor Each model for every instance run for a total 25 times to reduce the effect of randomness. SA was selected based on its performance in previous research (Martin 2013), as did the values for the cooling schedule.

The tests were performed in an HP EliteBook with 8GB ram, 4x Intel i7-3540M processor Each model for every instance run for a total 50 times to reduce the effect of randomness.

3.5 Conclusion

Although there is some research in fairness and social welfare measures in the literature in both applications and theoretical work a complete investigation including most measures is yet to be done. A number of measures from economics and social sciences have never been applied in OR. This could be done in one or more OR problems.

In case that this investigation has been done (Vasupongayya and Chiang 2005, Bertsimas et al 2011a, Ouelhadj et al 2012) it is done with a few measures, or measures from same set of functions, and there is no valid way of evaluating the results of the objective functions that are based on those measures. In order for the outcomes to be reliable a number of different problems must be examined.

In this chapter we presented a framework to evaluate potential inequality measures. A list of measures was presented and evaluated according to that framework.

Chapter 4 Methodology

4.1 Introduction

The goal of this chapter is to present recurrent themes that appear in the next chapters. Information about the Nurse Rostering Problem and the benchmarks used will be provided. The search method used in experiments and details about it will be presented. Methods to evaluate and compare results produced by measures such as Lorenz curves (Lorenz 1905) ELECTRE method (Roy 1968) and a variation of ELECTRE will be introduced.

4.2 Nurse Rostering Problem

The majority of research on NRP has been with the goal of minimising the sum of soft constraint violations (Burke et al., 2001). In this work, the goal is to improve nurses' satisfaction by either improving individual schedules, or eliminating the existing inequalities that can occur. This is done in the context of happiness satisfaction.

4.2.1 Benchmarks Used

In order to evaluate the quality of nurse rosters produced by the measures they were applied in benchmarks from Bilgin (2008). Those benchmarks provided 4 different wards with 2 different scenarios: emergency, geriatrics, psychiatry and reception wards, with nurses having identical or variable preferences. The wards vary in parameters such as number of nurses needed, scheduling period, number of shifts and skills categorisation. The total number of constrains differ in each ward, with geriatrics ward being the most trivial and Emergency being the most complicated one. In the cases of scenarios with nurses having variable preferences, the constraints are personalised, but there still is a contract that is common for a set of nurses.

Instance	Nr of nurses	Nr of shifts	Nr of skills	Planning period
Emergency	27	27	4	28 days
Geriatrics	21	9	2	28 days
Psychiatry	19	14	3	31 days
Reception	19	19	4	42 days

Table 4-1: Hospital Wards in Bilgin

The model was designed as such. A number of skilled nurses were to be assigned to a number of shifts. Those assignments must satisfy some constraints. Those constraints could either be hard constraints that must be satisfied in order for the final roster to be acceptable, or soft

constraints, where when not satisfied a penalty is added in the objective function. Soft and hard constraints are given in table 4.3. Apart from nurse satisfaction, hospital coverage constraints are to be satisfied. That means that all shifts must be covered adequately, and with nurses that are skilled appropriately.

The mathematical model of nurse rostering has been given in Burke 2008 and expanded in Smet 2014.

Notations
I: set of nurses
sk: set of skill types
I_{sk}: set of nurses with a skill type
D: set of working days
S: Set of shift types
N_{DSsk}: Number of nurses with skill sk per shift in a day needed
R_{IDS}: Nurses absence requests
R'_{IDS}: Nurses working requests
I_w: nurses working hours
I_{st}: maximum desired consecutive days of working s shift for nurse I
I_{sh}: desirable upper bound of hours in shift type

Table 4-2: Nurse rostering Notation

Nurse Rostering	
Hard Constrains	Soft Constrains
• A nurse can be assigned to maximum one shift each day	• Coverage constrains
• No consecutive shifts	• Rest times honoured
• Nurses and shift skill type match	• Series Constrains
	• Counter Constrains
	• Multi skilled nurses
	• Successive series

Table 4-3: Soft and hard constraints

The constraints can be of different types: Single constraints, counters, series, successive counters or successive series. Table 4.4 presents the different constraints and what type they are.

Single constraint	Counters	Series	Successive Series
Request (absence/work)	Days Idle	Days Idle	Shift Idle/ Shift Worked
Collaboration	Days Worked	Days Worked	Days Idle / Days Worked
Training	Hours Worked	Shift Types Worked	Shifts Type Worked / Days Idle
	Shift types worked	Weekends Idle	Days Idle / Days Worked
	Weekends Idle	Weekends Worked	
	Weekends worked		

Table 4-4Categorisation of constraints

Single constraints restrict the occurrence of specific events such as requests on working on a particular day/shift or request for absence for a particular day or shift.

Restrictions on the number of incidents of specific subject in a limited period of time are called Counters. An example could be the number of total working hours within a week.

Restrictions on the consecutive number of incidents of a specific subject in a limited period of time are called series. An example could be the maximum number of consecutive working days.

Restrictions on the consequent number of incidents of a series of specific subjects in a limited period of time are called successive series. An example could be the number of days off after a period of working night shifts.

Some examples of constraints are in the table 4.5, with bold being the hard constraints.

Constraint	Description
$$\sum_{d}\sum_{s} s_h * I_{sd} + l_1 - l_1' = 0$$	Workload constraint
$$\sum I_{D-1} - I_D + I_{D+1} + l_2 \geq 0$$	Limit on Standalone shifts
$$\sum R_{IDS} - l_3 = 0$$	Absence request Satisfaction
$$\sum R'_{IDS} - l_4 = 0$$	**Shift request satisfaction**
$$\sum_{I}\sum_{d}\sum_{s} I_s - l_5 \leq I_{sh}$$	Counter Restriction of (undesired) shifts
$$\sum_{I}\sum_{d-d+5}\sum_{s} I_{sd} - l_6 \leq I_{st}$$	Series of undesired shift
$$\sum_{I}\sum_{d=6/7,13/14\ldots}\sum_{s} I_s - l_7 \leq 2$$	Working in weekends
$$\sum_{I}\sum_{d,}\sum_{s} \lvert I_{sd} - I_{sd+1} \rvert - l_8 = 0 \, for \, d = 6,13,20,27$$	Working both days in weekend
$$\sum_{I} D + I_8 \leq 1$$	**Working maximum 1 shift per day**
$$\sum_{d-d+1}\sum_{s} I_{ds_3} + I_{d+1s_1}$$ $$+ \sum_{d-d+1}\sum_{s} I_{ds_i} + I_{d+1s_{i+1}} - l_9 \leq 0$$	**Number of consecutive shifts**

Table 4-5Nurse rostering constraints

The objective function for all models is a function of coverage constraints and nurse satisfaction constraints.

$$x = \sum_{l=0}^{n} w_i * l_i$$

$$cov = \sum_{icov=0}^{m} w_{icov} * l_{icov}$$

$$of = F(x, cov)$$

Where x are the weighted sum of constraint violations l_i. cov is the sum of weighted constraint violations that are for coverage and of the objective function used in the next chapters.

4.3 Lorenz Curves

In the following chapters a variety of different measures are going to be used as objective functions. Those measures are assumptions on the view of the decision maker, and are tailored to the decision makers' preferences in each case. However, in the process of deciding what measure leads to better results some other methods are needed, while making the least amount of assumptions possible. Those methods are Lorenz curves, and ELECTRE method.

Those methods can't be used as objective functions. Both of them can lead to incomparable results, that would render the search process impossible. However, they can be used to compare the results produced by other methods.

4.3.1 Lorenz curve

Lorenz curve, introduced by Lorenz (1905) is a way of visualising distributions. It shows the cumulative proportion of sorted incomes. It focuses just on equality between individuals.

Lorenz curve calculates the cumulative distribution of an allocation. It can be created by the following steps: First, the utilities must be increasingly sorted. Then the percentage of utility for every individual is computed. The last step is to figure the cumulative percentage of utility.

Every distribution creates a monotonous curve with the first node at (0.0) and the last at (1.1). If the allocation is completely fair, then the curve would be a straight line (x=y) from 0 to 1. The closer a curve is to the straight line, the fairer the allocation is. If distribution A is completely below (since deal with a minimisation problem) distribution B, then the allocation A is considered to be more equal than allocation B, since it is more equal for every quantile. However, if the curves are crossed no conclusion can be made. An example of Lorenz curve can be shown in figure 4.1. In this case, it can be seen that the results produced by "average" are less equal while results produced by "Atkinson 1.9" appear to be more equal than the others.
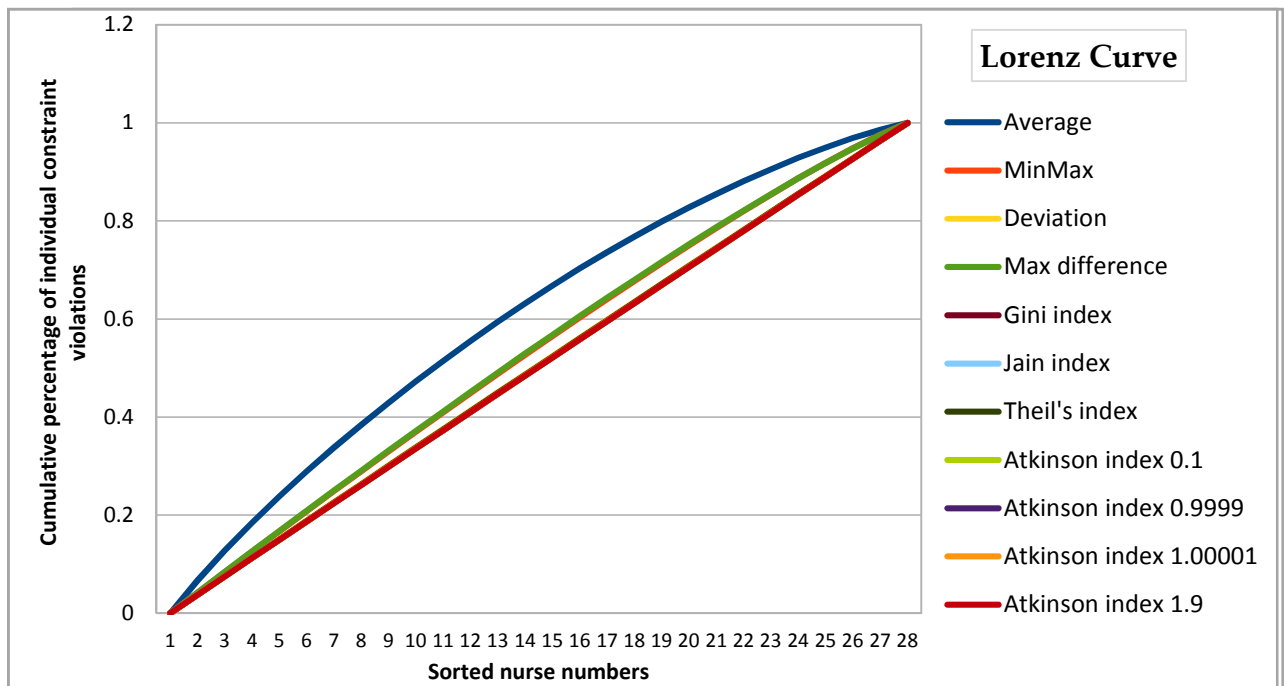
Figure 4.4-1: Inequality Lorenz Curve

4.3.2 Generalised Lorenz Curve

To have a more complete evaluation of a distribution, Lorenz curves are not enough since they don't incorporate an efficiency aspect. In order to incorporate efficiency (or welfare), Generalised Lorenz (GL) curves (Shorrocks 1983) must be used. GL curves are defined similarly to Lorenz curves: Again, the individuals are sorted in descending order. In both Lorenz and GL curves X axis represents the cumulative percentage of the population. In Lorenz curve Y axis represents cumulative percentage income while in GL curves Y axis is cumulative income. While in Lorenz curve the range of values in the Y axis is from 0 to 1, in GL the curve is from 0 to the total value of constraint violations. An equal distribution would still be a straight line. A non-equal distribution would be considered better than an equal one if the non-equal is wholly below the equal one(in case of minimisation problems).Again, if lines are crossed, no conclusions can be drawn.

Generalised Lorenz curves are an interesting welfare tool: they account for both equality and effectiveness. An example of Generalised Lorenz curve can be shown in figure 4.2. In this example, while results produced by "average" appear to be less equal would be still preferable than all the other results, since the curve lies below all other curves.
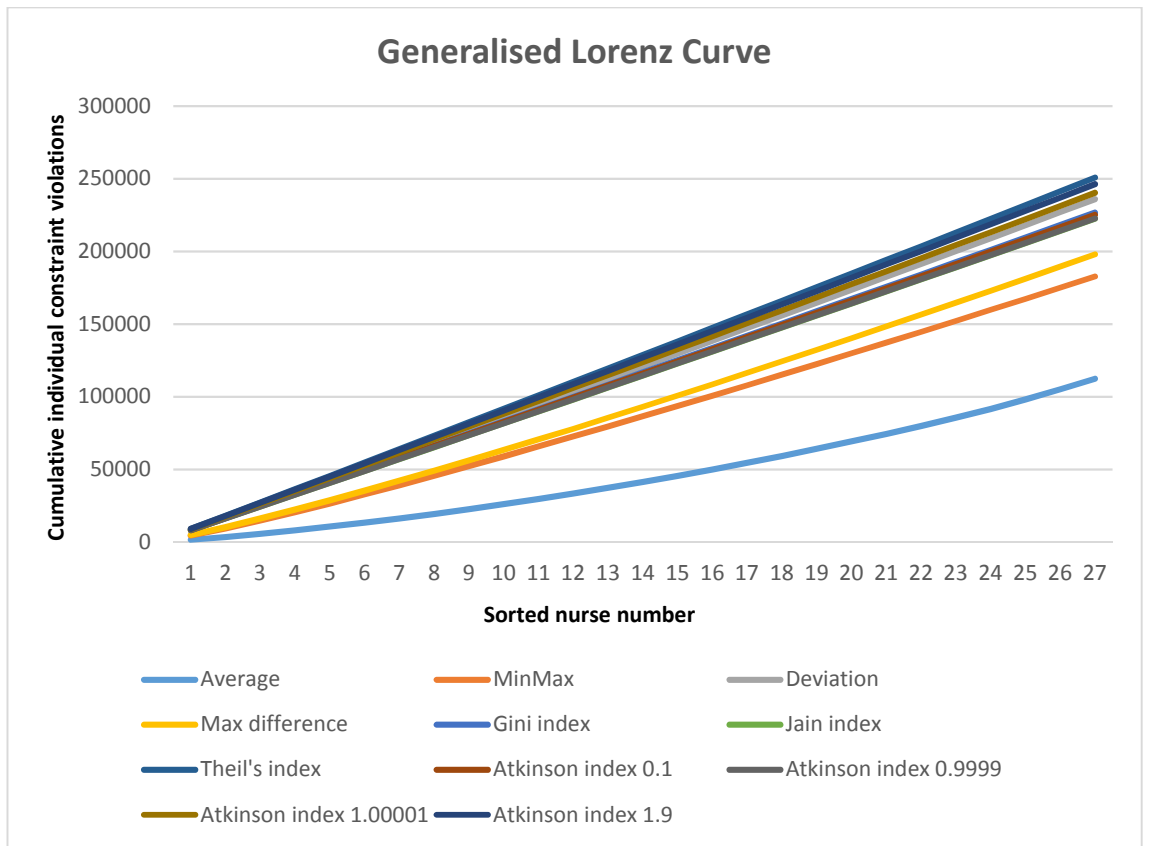
Figure 4.4-2: Fairness Generalised Lorenz Curve

4.4 ELECTRE method

Another method to evaluate results is by ELECTRE method (Roy 1991). ELECTRE's I goal is to develop and use an appropriate outranking relation between alternatives. The four possible scenarios considering two alternatives A and B are:

- Alternative A is preferred over alternative B
- Alternative B is preferred over alternative A
- Alternative A is indifferent to alternative B
- Alternative A is incomparable to alternative B

To produce relationship, there are two main steps: The concordance, and the disconcordance test.

The concordance test is the evaluation of the evidence supporting the claim that one of the alternatives is dominating the other. This is achieved by checking the performance of A and B in the individual criteria.

$$C(A,B) = \sum_F w_i$$

Where i is the criterion, F is the set where $A_i \geq B_i$ and $C(A,B)$ is the concordance-evidence that alternative A is better than alternative B. Values of $C(A,B)$ range from 0 to 1. The higher the value, the stronger the evidence that supports $C(A,B)$. In order to formulate concordance a threshold s must be defined. If $C(A,B) \geq s$ that means that there is concrete evidence that alternative A is better than alternative B.

While concordance test clarifies the dominance of alternative over the other there might be some criteria that are critical to form dominance. If in those specific criteria the evidence is contrary to the ones obtained in the concordance test, a veto condition is imposed.

$$V_i(A,B) = \begin{cases} 0 \ if \ A_i - B_i < t_i \\ 1 \ if \ A_i - B_i \geq t_i \end{cases}$$

Where A,B the two alternatives, V_i is the veto for criterion i, A_i, B_i the performance of A,B alternative in criterion i, and t_i the veto threshold for criterion i.

The scenarios described above are constructed as such:

- Dominance of A over B if $C(A,B) \geq s$, $C(B,A) < s$ and $\sum v_i(A,B) = 0$.
- Dominance of B over A if $C(B,A) \geq s$, $C(A,B) < s$ $\sum v_i(B,A) = 0$.
- Indifference between A and B if $C(A,B) < s$ and $C(B,A) < s$ or if $C(A,B) \geq s$ and $C(B,A) \geq s$.
- Incomparability if $C(A,B) \geq s$, $C(B,A) < s$ and $\sum v_i(A,B) \neq 0$ or if A if $C(B,A) \geq s$, $C(A,B) < s$ $\sum v_i(B,A) \neq 0$

In this application of ELECTRE I, the alternatives are the results produced by each model and the criteria are each nurses' sum of weighted constraint violations. For each nurse i:

$A_i = \sum_{j \in C} w_j$

Where c is the set of confirmed constraint violations for this particular nurse, w_j is the weight constraint violation j. So A_i is represented as the value of criterion i for alternative A in ELECTRE I terms.

In order to use the ELECTRE I method the nurses are sorted by sum of weighted constraint violations, beginning from the ones with most constraint violations to the ones with least. This is needed for the weight determination (see 4.4.1) and reflects fairness: the ones with more constraint violations are the ones worse-off.

In order to obtain concordance, the same formula would be used:

$C(A, B) = \sum_F w_j$ where

$\quad$ A, B are the two different results obtained from different measures.

$\quad$ w_j is the sum of weight, given for each nurse, based on the sorting position.

$\quad$ F is the set where $A_i > B_i$, the set where the sorted "nurse" from result A, has less weighted constraint violations from the sorted "nurse" from result B. :

$$A_i = \sum_{k \in C} w_{A_k} \leq B_j = \sum_{k \in C} w_{B_k}$$

w_{A_k} and w_{B_k} respectively are the weights for the constraint violations for the sorted "nurse" i for solution A, B.

In order to produce the veto results, again the sum of weighted constraint violations is considered to be a nurse. In the use of ELECTRE I in next chapters (8.4.3, 9.5.3) the criteria-nurses used for veto was a percentage of the nurses with the highest sum of weighted constraint violations. A percentage was chosen because it would be expected in different populations to have different number of those considered "poor" or able to impose veto. The threshold was set as a factor of the performance of the other alternative's (roster) criterion (nurse) in the specific criterion:

$$\frac{A_i}{B_i} \geq t$$

This threshold ensures that in a number of sorted "nurses" the performance of solution A is not more than t times worse than the performance of solution B for those specific sorted "nurses".

4.4.1   Fair ELECTRE method

An issue that can be risen with ELECTRE method is the importance given on each criterion. Since the focus is mainly inequality concepts, the weights must be arranged accordingly. Determining weights is a complex procedure and can result to differences in evaluations.

In order to incorporate equality concepts, the weight for the most deprived individuals should be increasingly higher than the least deprived individuals. The formula for the weights in the ELECTRE method was this:

$$wt_i = \frac{(K - r_i + 1)^z}{\sum_{j=1}^{K}(K - j + 1)^z}$$

Where K is the number of nurses, r_i is the sorted position of the nurse the weight is going to be assigned, j is the sorted position of nurses, and z is an equality factor. The higher z it is, the more inequality averse ELECTRE is. The concept of z as providing inequality averse viewpoints resembles the one from Atkinson 1971. This assignment of weights was introduced in Stillwell 1981, and while it has been used in multiple criteria decision making (Barron and Barrett 1996 Alfares and Duffuaa 2008), has not been used in fairness context. This formula can lead to a range of different inequality perspectives. If $z = 0$ the weights for each criterion would be the same leading to a utilitarian context. If $z \to \infty$ the most important criterion is given 100% of the weight.

4.5 Conclusions

In this chapter recurring themes that appear this research were presented. Nurse rostering problem was described. Methods to evaluate and compare results were presented. Those methods are creating a framework of comparing results produced by different objective functions, with minimum assumptions made.

Chapter 5 Poverty concepts

5.1 Introduction

Another way of measuring allocations quality is the concept of poverty. In this case, individual utilities are compared with a threshold, the poverty line. Poverty line is a utility line that if an individual is below it, he can be considered poor. Being starved of utility, whether that translates to actual income, resources, preferences or some kind of goal satisfaction leads to entities that are not only dissatisfied, but also unable to function.

Poverty consists of two different stages. The first one is identifying the individuals that can be described as poor. The second is to compare in which distribution there is more poverty between two different ones. Thus, the first one can be identified as "Identifying Poverty" and the second one as "Measurement of Poverty" can be described as a function of individual Poverty.

As far the first question is concerned, there are a series of work by, among others, Rowntree 1901, Townsend 1954 and Atkinson 1970. Defining poverty and the groups of people that can be considered poor is a major part of poverty research in economics. In brief and since it is not a part of this research, criteria for determining poverty can include means of sustenance, basic needs (clean water, housing etc) and clean environment. Size of a family is considered as well. Work has been made to take into consideration cost of living and the differences between countries or regions. The most dominant one is actual income. However, the principal and most interesting feature is the definition and outline of "poverty line".

Poverty line is a threshold which below that an individual is starved. With a poverty line defining a series of measures can be arise. The usual targets of the measures are to minimise the number and the intensity of the effects of those under the poverty line. However, in order to do that there are a number of ways or measures to evaluate poverty, some of them shown on section 5.5.

The Poverty concept is different than inequality in a series of ways. The main difference is that in equality individuals are compared amongst them, while in poverty the main comparison is with individuals and poverty line. An inequality averse scheme tries to minimize differences between individual entities in a distribution. A poverty aversion scheme generally targets individuals that do not have enough utility to satisfy their needs. While

targeting inequality, an equal distribution may be reached while having unhappy individuals that do not have enough utility to satisfy their needs. Accordingly, when targeting poverty a distribution where everyone is happy -satisfying their needs – while being completely unequal might occur.

Poverty presents with even more applicable issues than inequality and fairness methodology. The evaluation can again be either ordinal (Sen 1976) or under certain assumptions cardinal. Accordingly fairness measures are only interested in cardinal measures. The choice of the measures or of what is constituted as poverty can be subject to personal views. Also, the choice of the poverty line can lead to different evaluations (Foster and Shorrocks 1988b). Defining the poverty line itself is a matter of discussion. This creates a series of issues that make the use of poverty measures almost non-usable in many OR problems, unless a number of assumptions are imposed. Another concern is that there is no generally accepted terminology, leading to different measures with the same name (see Beckerman 1979 and Clark 1981). Also, Poverty measures are non-convex, that hamper the search procedure.

5.2 Poverty concepts in economics and OR

Poverty concept is originated from economics. Among the first mentioning poverty was A. Smith (1776), where emphasis is given at the changes of what is constituted poverty depending on the practices and the means of production on the era and the abundance of products. Karl Marx (1867) mentions that "in a given country, at a given period, the average quantity of the means of subsistence necessary for the labourer is practically known" implying an absolute poverty line.

Booth (1889, 1891) was the first one to study poverty needs, and setting a numerical poverty line. An important aspect of his study was to survey the income and occupation of common people in London. Using the data from his survey Booth created a map that illustrated poverty rates in neighbourhoods.

Frankfurt (1987) focuses on the premise that having "enough" is an adequate condition for equity. He discriminates between different views of having enough, from having sufficient

utility to survive to being satisfied with a state of affairs. Apart from connecting equity to poverty, these views on morality also imply differences on poverty levels.

5.2.1 Poverty in applications

The vast applications of poverty concepts are in the field of economics. A series of studies have been made to either evaluate a poverty line (Kilpatrick 1973, Rainwater 1974, Orshansky 1965, Rowntree 1901 Goedhart et al 1977) or to study poverty in a specific demographic area (Anand 1977 for Malaysia , Tsakloglou 1990 for Greece, Demery and Squire 1996 for Malaysia again, to name a few). However, in all cases the use of poverty is descriptive, where poverty through the years is shown and preferences are depicted.

Kilpatrick (1973) sugests that poverty line raises when average income raises, using evidence from Gallup poll. Orshansky (1965) researched the poor between different categories (sex, race age, etc) and suggested that since we are dealing with varied poor population, a variety in measures is needed to reduce it, ending with the poetic "The poor have been counted many times. It remains now to count tl1e ways by which to help them gain a new identity." Goedhart (1977) calculated the poverty line based on subjective opinions on what is considered poor.

The concept of starvation of resources is applicable in a number of OR problems. The Lock/deadlock problems in Computer science, where processes that need a specific resource are blocked as a concequence of that resource is being used by another process (Holt 1972). This results to case where all processes are blocked and starved of resources. The solution for deadlock avoidance is usually algorithmic. The first solution for deadlock prevention was given by Dijkstra (1965).

Applications include diverse problems. Some of them are: Risk management (Roy 1952, Fishburn 1977), Economic cases such as Social policy and planning (e.g. Ferrera 2005) and lock cases (see 5.2.2.1) such as Network Resource allocation (Litzkow et al 1988) and Job shop scheduling (Eg Glassey and Resende 1988).

In Roy 1952, the safety-first principle denotes that the gross returns should be no less than the expected returns. According to Atkinson (1970) this corresponds to the headcount poverty measure. Ferrera (2005) analyses southern European countries' reasons behind poverty, those countries strategy fighting poverty and welfare state and compares them with other advanced countries (such as Scandinavian ones) using mainly headcount (see 5.6) and relative poverty (see 5.5).

Litzkow (1988) developed a system that identifies and prevents idle workstations. In Glassey (1988) a closed-loop mechanism was implemented in order to enforce "starvation avoidance". Queuing theory and inventory control concepts were adapted for a for stochastic Job shop scheduling problem.

However, while there might be a few instances that poverty concepts are utilised, it is not widely acknowledged as such. While in fairness there has been a few instances where it is recognised as a desired property, poverty has not been accepted at all. Even though there are cases (see 5.3) where poverty ideas have been used, this is generally coincidental.

5.2.2.1 Lock and deadlock cases

A similar concept with poverty or starvation of resources is lock, or deadlock in several logistics applications. For instance, in Gold (1978), deadlock is defined as "a situation where one or more concurrent processes in a system are blocked forever because the requests for resources by the processes can never be satisfied". Similar definitions were given In Moorthy et al (2003) "a situation where at least a part of a system stalls" and Kim et al. (2007 "a situation where one or more concurrent processes in a system are blocked forever because the requests for resources by the processes can never be satisfied".

Deadlock notions are present, apart from computing, in logistic problems. Manufacturing systems is a common application (Egbelu and Tanchoco 1984; Lim et al. 2003; de Koster et al. 2004; Le-Anh and de Koster 2005).

Deadlock handling for real-time control of automated guided vehicles (AGVs) at automated container terminals deadlocks occurring in automated seaport container terminals, where quay cranes unload containers from vessels and place them on AGVs. Deadlock prevention in automated guided vehicles (AGV) is one common application (Lehmann et al 2006, Moorthy et al 2003)

Most of the poverty concept problems, like the evaluation or the status of poverty line are being overridden by the nature of the problem. Poverty line is set as the demands of a node, and most of the times a solution is simply not being accepted if a starved member exists.

Even though the case of deadlocks and locks has similarities with extreme poverty, such as starvation of resources, critical methodological causational and solution approach differences exist. The "poverty line" is essentially the complete lack of available resources. In economics poverty is used as an evaluation tool. It is considered as something that is

undesirable but since poverty is used for evaluation rather than decision making, can be acceptable. When deadlock occurs in OR problems, most of the times the result is considered as an infeasible solution. The solution approach is always algorithmic: in most part of the research, work is being done to find the deadlock-free solution, not quantifying the effect a deadlock would cause. This leads to a simplification of the poverty concepts.

5.2.2 Summary – Connection with OR

Any problem that can incorporate fairness or satisfaction can possibly incorporate "minimisation of dissatisfaction" (also see 2.3). Poverty can be possibly used in the same sense that fairness can be used in all "societal" OR problems. This leaves space for further application of poverty concepts as evaluation and assessment of distributions.

Deadlocks have similarities with the poverty concept in the form of a general limited resources provided to entities. The main analogy is the ability to function on a satisfactory level when this luck of resources occurs. The difference between locks with poverty is that the former is an algorithmic concept, while the latter a decision making one.

However, there are OR problems, where resource-starvation free solutions would be an enrichment of the decision-making process. Several OR problems can utilise the poverty concept, including recourse allocation problems, scheduling (ie personnel staffing), facility location and the rest of societal OR problems.

MCDA problems also use utilise poverty concepts. Goal programming and the similarities between goals (see 5.3). Poverty can be identified by using multi criteria analysis, see Alkire 2010. In this sense, poverty is more connected with MCDA more than any other part of OR.

5.3 Poverty and goal programming

In order to further the discussion between poverty and Operational research an attempt to introduce and parallelise poverty concepts and MCDM and in particular possible similarities or applications between goal programming and poverty. For instance, the poverty concept is close to Goal programming (GP): In GP the concept of "target levels" that must be satisfied is being used. An analogy can be made between maximisation GP problems and the poverty gap. In GP, there are cases where minimisation of the underachievement of a target value must be performed. This is equivalent with poverty gap: Target levels of GP and poverty line must be reached. Namely, for a series of goals with target level s, a goal could be formatted as

$$\min a = \sum \frac{w_q n_q}{k_q}$$

s.t

$$f_q(x) + n_q = b_q, q \in N$$

Where $f_q(x)$ is the value of the alternative in the q criterion, n_q the negative deviational variable or the under-achievement of a goal w_q is the weight of goal q and b_q is the goal value.

Considering the factor $f_q(x)$ to be the income of an individual, b_q as the poverty line, and the negative deviation variable n_q as the poverty gap, the most common poverty measure can be constructed as such:

$$G_w = \sum_{i=0}^{i=n} w_i(z - y_i) \text{for every } u_i \leq z$$

The above measure is weighted poverty gap, which appears in economics when household level data are being used providing more importance in larger families.

The differences arise when the objective function is constructed. Both GP poverty measures use the deviation from a goal, but the aggregation of those deviations is different. In GP apart from the weighted sum the other two main variants (lexicographic and Chebyshev) are combining the goals with a priority concept. In Lexicographic GP a number of priority levels exist. In each level there is a number of unwanted deviations to be minimised. In Chebyshev (or Minmax) GP the maximal deviation from any goal is taken into consideration(see Jones and Tamiz 2010). When using poverty measures, a numerical function is constructed on the deviation of the poverty line is the only way to evaluate things, contrary to Chebyshev/lexicographic approaches of GP.

While the simplest approach on defining poverty is to compare income with a poverty threshold, more indicators can be also used. For instance, Ravallion (1996) argued that four not purely income indicators can be used to measure poverty.

Human development index, introduced as a term by United nations in 1990, in a project lead by Haq (1990). Bourgignon (2003) used processes to combine the information of different monetary and non-monetary attributes in order to define poor members of society. Multi dimension poverty, introduced at 2009 by Alkire et al combines three different poverty "dimensions" - criteria and their "indicators" –sub criteria in a weighted program with

identical weights for each dimension and indicator in order to identify individual poverty. This method is used by the United Nations Development Program to date.

Those approaches resemble MCDA techniques, combining a set of Criteria into an evaluation. These techniques can be used not only in measuring poverty, but also identifying poverty. This creates an opportunity for a discussion, on the possibility of employing other functions in the context of GP other than weighted sum, like Watts index or Sen measure. The trade-off between complexity, accuracy and practical use could be a matter of debate.

For example, using the squared poverty gap instead of weighted GP, the problem could be formulated as such:

$$\text{mina} = \sum \frac{w_q n_q^2}{k_q}$$

s.t

$$f_q(x) + n_q = b_q, q \in N$$

5.4 Poverty Framework

There are a series of measures used in the literature ranging in complexity. The simplest one (headcount) enumerates the number of individuals below the poverty line, and more complex ones even incorporate fairness measures such as Gini.

In this section the poverty measures used in the experiments, some general desired axioms are going to described, and the aforementioned measures according to those measures are going to be analysed.

To begin with, the population of a distribution as a vector $x = (x_1, x_2, .. x_n)$ is defined. Without loss of generality, it can be assumed that the elements of x are sorted. For any poverty line z, the individuals can be identified as poor or not poor. There are two definitions of poverty, weak and strong poverty. The difference is in the inclusion of individuals exactly on the poverty line in the set of poor people or not. In the weak definition, if $x_i < z$, x_i is considered poor. In strong definition, if $x_i \leq z$, x_i is considered poor. For future reference, unless defined otherwise it is assumed strong definition of poverty.

Like the fairness measures, there have been some efforts to create a framework for poverty measures. Beginning with Sen (1976) set up some desired axioms to be used in poverty measurement. Foster et al(1984) added to Sen's axioms in order to create a class of poverty indices. A more complete framework was formulated by Zheng (1997), including desired and needed axioms for poverty measures. Alkire et al (2009, 2010) redefined the proposed framework to fit the concept of multidimensional framework. The main difference between desired properties in fairness and poverty is the existence of poverty line. For example, transfer principle in fairness examines if the difference between individuals is reduced, while in poverty generally if the difference between poverty line and an individual is reduced. However, there are many properties that are identical (Eg, anonymity, computational feasibility, principle of population).

In this framework, a distinction between core properties and some desired ones can be made. The core properties are:

- Focus axiom

- Anonymity (Symmetry)

- Principle of population

- Continuity

- Poverty line property

- Transfer principle

- Subgroup consistency

- Monotonicity property

- Computational feasibility

All of those properties – even with different names -were proposed in Zheng (1997) except from computational feasibility. This is to be expected since the needs in OR and Economics are different. Feasible computational cost is more needed in optimisation than description: in optimisation a poverty measure must be calculated a number of times, while in description only one.

The focus axiom ensures that the entities on question are the ones below the poverty line. It implies that distribution between non-impoverished members is irrelevant to the

evaluation of the distribution. Thus, an increase on the utility of a non-poor person should not affect the total evaluation. This also implies that if more information for the total evaluation of a distribution is wanted some other measures or concepts must be utilised.

Even though non-poor members are not used in the evaluation of the distribution, information from them can be useful in other ways in the concept of poverty. An example could be the estimation of difficulty in eliminating poverty by transferring utility from the non-poor to the poor. However, this is beyond the purpose of this research.

Anonymity ensures that no one is discriminated against by any attribute such as name, race or ID. The same concept is used in this inequality framework.

Principle of population is also required to be applied. It ensures that if a distribution can be replicated to a different size distribution with the same distribution, their poverty will remain the same. The same concept is used in this inequality framework.

Continuity implies that every change on an individual below (or in the case of not restricted continuity, on) the poverty line would result on a change of poverty levels. The poverty levels are becoming progressively more severe the lower utility an individual has (Watts 1968). Also, small changes in an individual utility should not lead to a huge change to overall poverty. However, a poverty measure is not possible to be continuous on the poverty line when increased. A change on an individual from poor to nonpoor would necessary lead to a change on overall poverty.

A change in poverty line could lead to changes to poverty, all other things equal. Zheng (1997) states that "increasing a poverty line would lead to higher poverty level". Even when there is no change in status in any individual, increasing distance from poverty line would increase overall poverty.

Transfer principle in Poverty frameworks is slightly more complicated than in inequality. Donaldson and Weymark (1986) identified four different cases in transfer principle, sorted by the weakest case to the strongest. The main point of those four transfer case axioms is that a transfer from a richer to a poorer individual should decrease overall poverty, while a transfer from a poorer to a richer individual should increase it.

- Minimal transfer:

Poverty decreases when there is a transfer from the richer to a poorer individual, when both individuals were, and remain below the poverty line after the transfer.

- Weak Transfer

Poverty decreases when there is a transfer from the richer to poorer individual, while the richer individual could be either above, or below the line. The number of poor individuals should not change.

- Regressive Transfer

Poverty increases after a transfer from a poorer individual to a richer one with at least the poorer person is below the poverty line

- Progressive Transfer

Poverty decreases when a transfer from a poorer individual to a richer one with at least the poorer person is below the poverty line.

Adopting a transfer axiom is a matter of debate. Sen (1976) suggested the use of regressive transfer axiom. However, since Sen's poverty measure could violate it2, so Sen (1981, 1982) considered minimal transfer axiom as needed in a poverty framework, and regressive transfer as a desired property. Donaldson and Weymark (1986) proved that any poverty measure satisfying continuity and weak transfer would also satisfy regressive and progressive transfer axioms.

Subgroup consistency axiom was discussed by Foster and Shorrocks (1991). It implies that if poverty changes in a subgroup of a distribution, while remaining the same in the rest of the distribution, overall poverty should change as well. It is analogous to Monotonicity axiom: Monotonicity examines an individual while subgroup consistency groups within a distribution. In that point of view, it has ties to decomposability, in the sense that groups are involved.

Monotonicity again can be defined as weak or strong Monotonicity. Strong Monotonicity axiom states that a poverty index decreases whenever the utility of a poor individual increases. Weak Monotonicity axiom states that poverty decreases when the utility of a poor individual increases as long as this individual remains below the poverty line.

By computational feasibility it is implied that the number of computational steps that are needed in order to evaluate a poverty index are reasonably small. In economic cases studies

---

[2] In a case where two individuals are poor, and there is a transfer between the poorer one to the richer one that levitates the richer one above the poverty line could be indicative that overall poverty has decreased.

there is usually a sample of entities that are taken into consideration (Foster et al 1984, Anand 1977), so calculating the value of an index is relatively trivial. However, when a poverty measure is used as an objective function, it can be expected that it will be executed a number of times, and depending on the

Some desired properties could be:

- Scale invariance/Translation invariance

- Decomposability

- Monotonicity sensitivity

- Normalisation

- Poverty/non poverty Growth

Decomposability in poverty measures is similar to inequality measures. A decomposable poverty measure can merge the poverty index of a number of subsets to one, usually by adding them.

Monotonicity sensitivity implies that a poverty measure should be more sensitive in the lower spectre of utility. It is the equivalent of Dalton principle: transferring utility from the richer to poorer entities will reduce general poverty.

Scale and translation invariance in poverty measurement are again similar to inequality measurement. A poverty measure is scale invariant (relative) if there is not difference between a distribution, and this distribution scaled (multiplied). Translation invariant (absolute) is a measure that does not distinguish between a distribution, and a distribution that occurs by adding utility in every individual.

Like inequality measures, Scale invariance is preferred: In real world economy, all things equal, it does not matter if one use Pounds, Euros, or Mexican pesos. Poverty line and all incomes should be exchanged in the respective currency.

Normalisation in poverty resembles normalisation in inequality. Following focus axiom, if noone is below the poverty line poverty should be zero. Having an upper limit would facilitate analysis, like comparative analysis.

Poverty (non-poverty) growth states that every time an individual is added (or removed) from poor population, overall poverty will increase (or decrease). This crates complications with transfer axiom in some cases (Kundo and Smith 1983, Donaldson and Weymark 1986).

5.5 Poverty line

Poverty line issues are a matter of discussion since the introduction of poverty. The first mention of poverty line is on Booth (1889, 1891) and it was defined as half of a pound to a pound a week. Booth considered this amount as "a minimum amount necessary for a family ... to subsist". This arbitrary set of poverty line is a consistent feature to day. However, methods for choosing the poverty line have been proposed.

Choosing the value of poverty line can have an impact on results. Foster and Shorrocks (1988a) showed that poverty measures and results depend on the level of poverty lines.

Rawls (1971) implied a universal poverty line. Even though this works focus is not on poverty, but rather inequality, his target is "least advantageous people". He claims that "all persons with less than half of the median may be regarded as the least advantaged segment". This created a distinct trend of relative poverty that is continued to this day. Townsend (1974), Fuchs (1967) also proposed measures of central tendency like average and median. In EU legislation, the degrees of poverty are set as various degrees of the median equalised income (EU council of ministers 1975). Those poverty lines estimate the relative poverty within the population.

Having a percentage of an indicator as the poverty line can have a number of advantages and disadvantages. A universal poverty line requires a minimal knowledge of the problem and the situation. In the case of economics, a relative income poverty indicator can be applied universally, in all countries. Considering OR problems it can be applied to a number of different instances and problems.

Setting the poverty line as a percentage of the median could have some disadvantages. Accepting a specific percentage of an indicator (50% or 60%, or any other number) as poverty line is arbitrary. A poverty line that is a percentage on median could result to differences in poverty depending on the distributed utility; having a high (low) total utility/income would mean that the poverty line would be set high (low), without any difference in a primary characteristic such as real purchasing power (or in OR problems, real utility). Bradshaw and Mayhew (2011) present as an example the poverty in UK and Estonia: those countries both had the same poverty rate of 19% when the poverty threshold based on median in Estonia was set at 9770€ and in UK at 24380€. A poverty line based solely on income (or utility) without having some connections with a direct indicator lacks meaning: a poverty line should

represent a threshold that if crossed the individual would become deprived, unsatisfied or unhappy.

In nurse rostering context setting the poverty line as a percentage of an indicator could represent a couple of different things. As in countries, constraint violations might not have different values in different nurse rostering problems, or wards. Setting a relative poverty line would eliminate those issues. Another view could be that a nurses' threshold for happiness depends on the total happiness of the group.

Following Booths' concept of poverty line measured in absolute terms, United Nations and U.S.A have implemented standard-absolute poverty line, based on actual needs of survival. It is set to be in a 1.25$ at 2005 purchasing-power parity. This is a factual representation of deprivation threshold, based on needs.

However, needs can also be subjective, or hard to define:

A lot of research has been made to define poverty line in the field of economics. Atkinson (1970) treated poverty as a part of a political decision, where people were considered poor based on benefits rights. Kilpatrick (1973) and Rainwater (1974) determined a poverty line based on what the public opinion was on basic needs is using polls. Orshanksy (1968) and Rowntree (1901) set the poverty line based on what are the basic human surviving needs (housing, clean water, food etc). Goedhart et al(1977) introduced the Leyden poverty line method, a function based on survey questions. Among others, Atkinson (2003) Alkire et al (2009, 2010) proposed identifying the poor based on combining several areas of deprivation, resembling MCDA approaches. Specifically, in Alkire and Santos 2010 a set of 3 criteria (Health, Education, Standard of living) with 10 total sub-criteria were weighted in order to determine if a family would be considered poor or not (also see 5.3).

As in relative selection of poverty line, a same analogy can be made for an absolute value of poverty line. If sufficient research has been made to identify the impact of constraint violation and their tradeoffs, an exact value of total weighted constraint violations can be chosen. This leads to the same problems as with the poverty in the world, with an extra one: research for poverty line has been done for years from many economists in order to reach to a conclusion. It wouldn't be reasonable to expect the same effort to be made in a single operational research problem.

Rawlsian minmax measure can be a special reverse case of a poverty measure: What is the maximum chosen poverty line where there is no one that can be considered poor?

$$\max z \, , \text{where } u_i \geq z, \forall i$$

Poverty measures can be used with or complementary to inequality measures and social welfare: Social welfare can be again maximised, either subject to certain poverty levels, or used in a multi-objective way combined with poverty. This may include again a lexicographic approach or a weighted approach.

5.6 Measuring poverty

Watts (1968) defined poverty measures as "a function of individual measures and the poverty line". It indicates the intensity of poverty, or poverty level, in an allocation of a non-negative value to a set of individuals. Given a poverty line a poverty measure will evaluate a distribution with a numerical number. As in fairness, there is a diverse range of poverty measures used in the literature. A number of them are used in the following table (Table 1)

The poverty measures used are presented in the following table. Note that measures can be continuous or discrete. Atkinson (1987) defined poverty measures in a continuous fashion, while more common is the discrete way (Sen 1976 - 1979, Foster et al 1984, Zheng 1997, Watts 1968 etc).

5.7 Poverty measures

Head Count	$$H = \int_0^z f(Y)\,dY$$ $$H = \sum_{i=0}^{i=n} y_i, \begin{cases} y_i = 1 \ if \ y \le z \\ \quad y_i = 0 \ else \end{cases}$$
Head Count ratio	$$Hr = \frac{H}{N}$$
Poverty gap (Batchelder 1971) Other names include Deficit, Dollar gap[3]	$$G = \int_0^z (z-Y)f(Y)\,dY$$ $$G = \sum_{i=0}^{i=n} (z-y_i)\, for \ every \ y_i \le z$$
Squared poverty gap (poverty intensity)	$$SG = \sum_{i=0}^{i=n} (z-y_i)^{\wedge}2 \ for \ y \le z$$
Income gap ratio (Sen)	$$G = \sum_{i=0}^{i=n} (z-y_i)\, for \ every \ y \le z$$ $$I = G * H = \sum_{i=0}^{i=n} (z-y_i) * \sum_{i=0}^{i=n} y,\, for \ y_i \le z$$
Normalised deficit	$$D = \int_0^z \left[1-\frac{Y}{z}\right] f(Y)\,dY$$ $$D = \frac{1}{n}\sum_{i=0}^{i=n} 1-\frac{y_i}{z} \ for \ every \ y \le z$$
Watts Measure	$$W = -\int_0^z \log_e\left(\frac{Y}{z}\right) f(Y)\,dY$$ $$W = \sum_{i=0}^{i=n} \log_e\left(\frac{y_i}{z}\right),\, for \ every \ y_i \le z$$
Sen index	$$S = H_r\left(1-(1-I)\left\{1-G\left[\frac{H}{1+H}\right]\right\}\right)$$
Sen with Jain	$$S_j = H_r\left(1-(1-I)\left\{\frac{\left(\sum_{i=1}^{i=q} y_i\right)^2}{\sum_{i=1}^{i=q} y_i^2}\right\}\right)$$
Clark et al second measure:	$$1/c[1-(1-P^*)^c = \frac{1}{c}\sum_0^z \left(1-\left(\frac{Y}{z}\right)^c\right),$$ $$where \ c \le 1$$
Foster et al (FGT)	$$P_a = \sum_0^z \left(1-\frac{y_i}{z}\right)^a f(Y)\,dY,\, where \ a \ge 0$$

Table 5-1: Poverty measures

---

[3] Other versions might include normalised deficit

(on population - $G = \frac{1}{n}\sum_{i=0}^{i=n}(z-y_i)\, for \ every \ y_i \le z$) normalised (on poverty line-

$D = \frac{1}{z}\sum_{i=0}^{i=n}(z-y_i) \ for \ every \ y_i \le z$) or poverty gap ration (see below)

Some measures are given in both continuous and discrete forms. The reasoning behind this is while most OR problems, and nurse rostering in particular have a limited number of individuals those measures were proposed to be used in countries with millions of individuals in question.

Headcount/headcount ratio

Head count is the simplest and most widely spread poverty measure. It represents the number (percentage) of people that are below the poverty line. Because of its simplicity, it is the measure used by the U.S Census Bureau to determine poverty, and has been used in the past from United Kingdom, EU, and United nations as the measure of poverty. Headcount is also common in research on poverty (Demery and Demery 1991). However, since headcount only counts the number of deprived people, and ignores the severity of deprivation, using headcount ration might lead to misleading results.

Headcount does not satisfy most of the required properties, for instance most versions of the transfer axiom, or strong Monotonicity. Since it is not distribution sensitive, it also might violate poverty line axiom. Headcount is also non-continuous, and that can cause a series of issues when used.

For example, and in NR environment, suppose that there are two nurses with the same preferences and nurse A having more constraint violations than nurse B with both of them being unsatisfied from their schedule (ie, having more constraint violations than acceptable). If nurse's A shifts worsen (by getting more constraint violations) in the benefit of nurse's B roster (so that nurse B gets less constraint violations) and both nurses will remain unsatisfied Headcount will not detect this change.

Poverty gap

Poverty gap is the other widely used poverty measure, representing the total utility that must be distributed to the poor members of the society in order to elevate them from poverty. Being a relatively simple measure, it is widely used together with headcount ratio, for instance, in World Bank Data (World Bank Group 2012).

The income gap does not satisfy the weak transfer axiom, or strong monotonicity. Being a not distribution sensitive axiom does not satisfy monotonicity sensitivity.

Consider nurses A and B from previous example. If the amount of constraint violations traded between them is constant, weak transfer principle will not be satisfied:

$$Y = (y_1, y_2, \dots y_i, \dots y_j, \dots y_{n-1}, y_n)$$

Where n is the number of nurses, $y_1, y_2, \dots, y_n$ the constraint violation of nurses, z the maximum amount of accepted constraint violations, and y_i, y_j nurses that have more constraint violations than the ones accepted. If a transfer k happens between nurses i, j where

$Y = (y_1, y_2, \dots y_i - k, \dots y_{j+k,} \dots y_{n-1}, y_n)$ and $y_i - k > z$, the result of poverty gap

$$G = \sum_{i=0}^{i=n} (z - y_i) \, for \; every \; y_i > z$$

will remain the same.

Income gap (Sen)

This measure, proposed by Sen (1976) is the product between poverty gap and headcount. It has been used in research, for example in Anand (1977, 1983) and Thon (1979). However, even though it is a relatively simple measure, it has not been thoroughly used in poverty measurement in societies.

Being a product of poverty gap and headcount, Income gap violates all transfer and sensitivity axioms, while satisfies increasing poverty line and continuity, that headcount violates.

Zheng (1997) claims that this measure empirically can perform on an equivalent level as other "good" poverty measures.

Poverty gap squared

Poverty gap squared, or poverty severity index (Haughton and Khandker 2009) is the sum of the squares of the poverty gap relative to poverty lines. It is a special case of FGT index, (Foster et al, 1984). It has been used in research (Jalan and Ravallion 1979) and has been mentioned in international organisations (Haughton and Khandker 2009). However, it is not as common as the previous poverty measures.

Poverty gap squared does not satisfy weak transfer sensitivity, however at its ratio form satisfies all other properties.

Sen measure

Sen measure was introduced by A. Sen (1976). It is a measure that incorporates fairness, and as such is distributional sensitive. It can be calculated using the following formula:

$$S = H * \left(1 - (1 - G^p) * \frac{\mu^p}{z}\right)$$

Where H is the headcount index, G is the Gini index amongst the poor population, μ^p is the average amongst the poor population, and z is the poverty line.

Sens index is not continuous, subgroup consistent, or satisfying strong transfer principle. However, since it satisfies monotonicity and minimal/weak transfer, and it studies fairness within the body of poor individuals, it has been wildly discussed in academic context (eg Foster et al, 1984). However, Sen measure has not been used in practice. Lack of continuity or decomposability could be reasons for that (Deaton 1997)

Watts measure

Watts measure is the sum of the logs of the ratio of poverty line by income. Watts measure was the first distributional sensitive measure that was introduced (Watts 1968). It can be transformed to a function of Theils' index, as shown by Blackburn (1969)

Watts measure, even though it satisfies all properties, it has scarcely been used in the literature. A possible reason could be that logarithms are computationally expensive to compute.

Sen with Jain

This measure has the same concept as Sen measure, but instead of using Gini index, Jains (Jain et al 1984) index is used. It retains the inequality aversion and monotonicity sensitivity properties.

Jain's measure has not been combined with poverty before this work. The reasoning behind using Jain measure is that it has common properties with Gini index (bounded, transfer principle) and could be used as a measure instead of Gini.

Foster Greer Thorbeck

Foster Greer Thorbeck (FGT) was introduced by Foster at al (1984). It is a class of measures that, depending on the α-Factor, transforms to previously mentioned measures, or creates new ones.

The α-Factor sets the strength of distribution sensitivity: the higher it is set, the higher sensitivity the index has. For $a = 0$, FGT collapses to headcount, for $a = 1$, FGT equals to poverty gap, and for $a = 2$ FGT is the squared gap. For cases where $a > 2$, FGT satisfies all desired properties.

FGT has been studied in the literature (eg, Foster et al 1984, Atkinson 1987) but not thoroughly used in practice. An example of using FGT is Tsakloglou (1990)

5.8 Conclusions

In this chapter it is discussed the concept of poverty. While well documented in economics and social sciences, there has been very limited use of its tools in Operational research (OR), and those can only vaguely resemble the depth of analysis +that poverty concepts can provide in evaluating distributions.

A framework for poverty measures was setup, similar to Fairness concepts. A multitude of poverty measures were presented and analysed based on this framework and their use in economic research.

In this chapter the connection between poverty and OR applications is also presented. Poverty measures can be used as objective functions as they are, or in combination with other objectives, as a tool to evaluate distributions. This could add depth, and include the concept of minimisation of dissatisfaction in OR problems, where this concept can be fit.

Chapter 6 Inequality results

6.1 Introduction

While fairness measures have been used in practice in operational research (OR) (Bertsimas et al 2011, Martin et al 2012) there have been no efforts to evaluate the viability or perform an analysis of said measures. In order to investigate the properties of fairness measures, they were applied in Nurse Rostering problem. The goal of this chapter is to compare the proposed measures from literature with the existing measures from statistics.

6.2 Methodology

Since the goal of this chapter is to evaluate the performance of fairness measures coverage constraints were not taken into consideration. The rationale behind this is that fairness only results are not applicable since they overlook efficiency and general individual satisfaction (see 6.4).

The objective function was set as the fairness measure, calculated for nurses. As described in previous chapters, each nurse's sum of weighted constraint violations was considered an "individual" (y_i) (section 6.3).

6.3 Models

Since the Nurse rostering problem is a minimisation problem, using measures that are applicable in maximisation problems would lead to results conflicting to fairness goals.

While some measures can be the same to both minimisation and maximisation problems, like average or deviation, changes must be made to other measures in order to adapt.

For example, Jains index numerator and denominator are swapped. The logic behind this is that when trying to minimize an objective that would otherwise be desired to be maximised, minimising the inverse would be preferable.

Measure	Objective function
Average	$$WoOld = \sum_{i=1}^{i=N} y_i$$
MinMax	$FoaOld = max x_i$
Deviation	$$FobOld = \sum_{i=1}^{i=N} |y_i - \mu|$$
Max difference	$FocOld = max y_i - min y_i$
Gini index	$Foa = \dfrac{2\sum_{i=1}^{N} i*y_i}{n\sum_{i=1}^{N} y_i} - \dfrac{n+1}{n}$
Jain index	$Wo = \dfrac{\sum_{i=1}^{N}(y_i)^2}{(\sum_{i=1}^{N} y_i)^2}$
Theil's index	$$Fob = \frac{1}{N*\log(N)} \sum_{i=1}^{i=N} \frac{y_i}{\mu} * \log(\frac{y_i}{\mu})$$
Atkinson index 0.1	$$Foc = (\sum_{i=1}^{i=N} (\frac{\mu}{y_i})^{(0.9)})^{(\frac{1}{0.9})}$$
Atkinson index 0.9999	$$Fod = (\sum_{i=1}^{i=N} (\frac{\mu}{y_i})^{(0.000001)})^{(\frac{1}{0.000001})}$$
Atkinson index 1.00001	$$Foe = (\sum_{i=1}^{i=N} (\frac{\mu}{y_i})^{(-0.000001)})^{(\frac{1}{-0.000001})}$$
Atkinson index 1.9	$$Fof = (\sum_{i=1}^{i=N} (\frac{\mu}{y_i})^{(-0.9)})^{(\frac{1}{-0.9})}$$

Table 6-1: Fairness measures

Where y_i is the sum of constraint violations for nurse i, μ is the average constraint violations for nurses (or $WoOld$) and N is the number of nurses.

Theils' and Atkinson measure were also reversed. The results support those transformations.

Gini index, derived from Lorenz curves also changes. Lorenz curves in maximisation problems are below the equality line.

GINI index would be the area that is above the equality line and below the line of other allocations. The function that calculates it does not change substantially.

6.4 Evaluation

In this section the objective functions used were set to strictly minimise inequality. The rosters produced would not be accepted by any nurse, but the results are interesting in the scope of maximising fairness in an environment. The tool to compare equality between

individuals are Lorenz curves (Figure 6.1) (see 4.3.1). The measures and objective functions used were the ones presented in table 6.1.

Lorenz curves were also used to compare results (4.3). The adaptation of Lorenz curves in minimisation problems is that Lorenz curve lies above absolute equality instead of below it: in every possible allocation the individuals would have more, or at best the same, constraint violations than the constraints violations in individuals in the most equal allocation. Again, like maximisation problems if two curves are crossed no dominance can be established.

The Lorenz curves are grouped into 3 distinct groups. Average is the least equal between them as expected. The curves that are somewhat closer to the equality line are the ones produced by Minmax and Max difference. All the other results were really close to equality line, but still some of their Lorenz Curves were Crossed. The result from "Atkinson index 0.1" provided the results that were the least inequalities, according to Lorenz Curves since the curve produced by it was the one closer to the equality line at all times without crossing with any other. The results produced by "Atkinson index 1.9" were the second most equal, again without crossing with any other line.

The Lorenz curves of Atkinson 0.99999, 1.00001 and Gini were crossed. In the six worse off nurses, Atkinson 1.00001 was better (lower). Then at the point the 7th nurse the curve for Atkinson 0.99999 was lower than the others. At the point of 8th to the 13th nurse Gini curve was preferred. The rest of the curve saw a couple of crossings between Gini and Atkinson 1.00001, making the results below the two first incomparable.
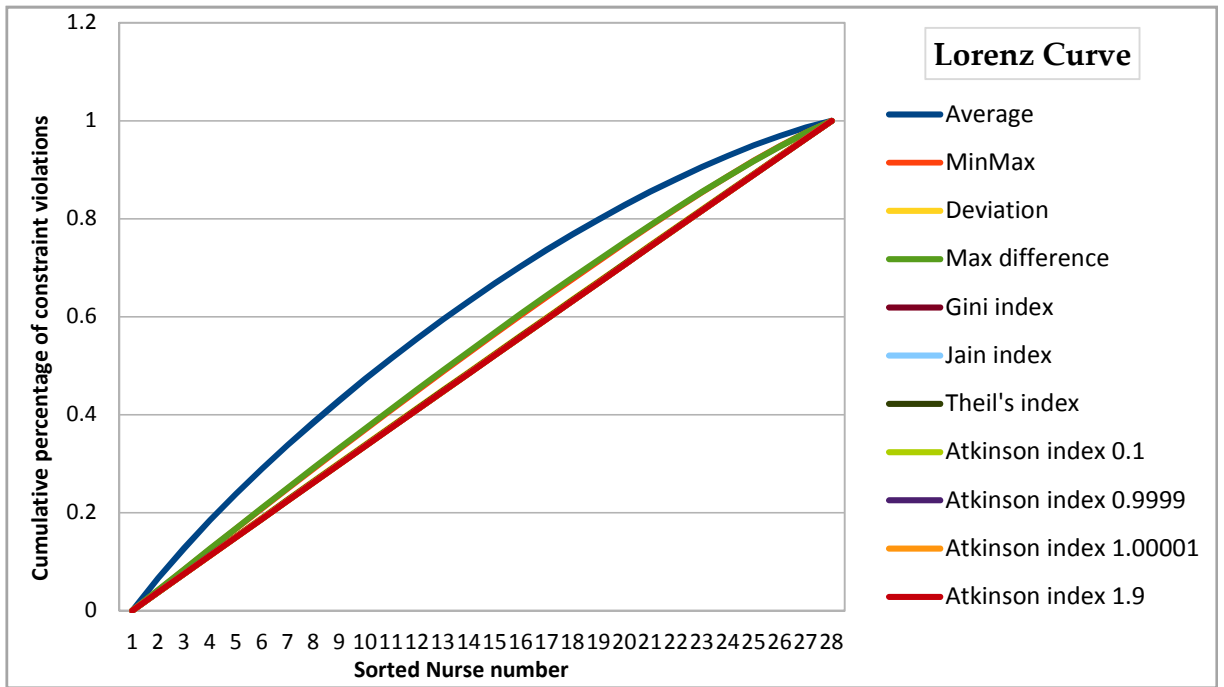
Figure 6-1: Inequality Lorenz Curve

Some evaluation must be made for efficiency, using Generalised Lorenz curves. Contrary to the equality results, Average had the best performance, being lower than the other curves. Then the MinMax and Max difference measures followed.
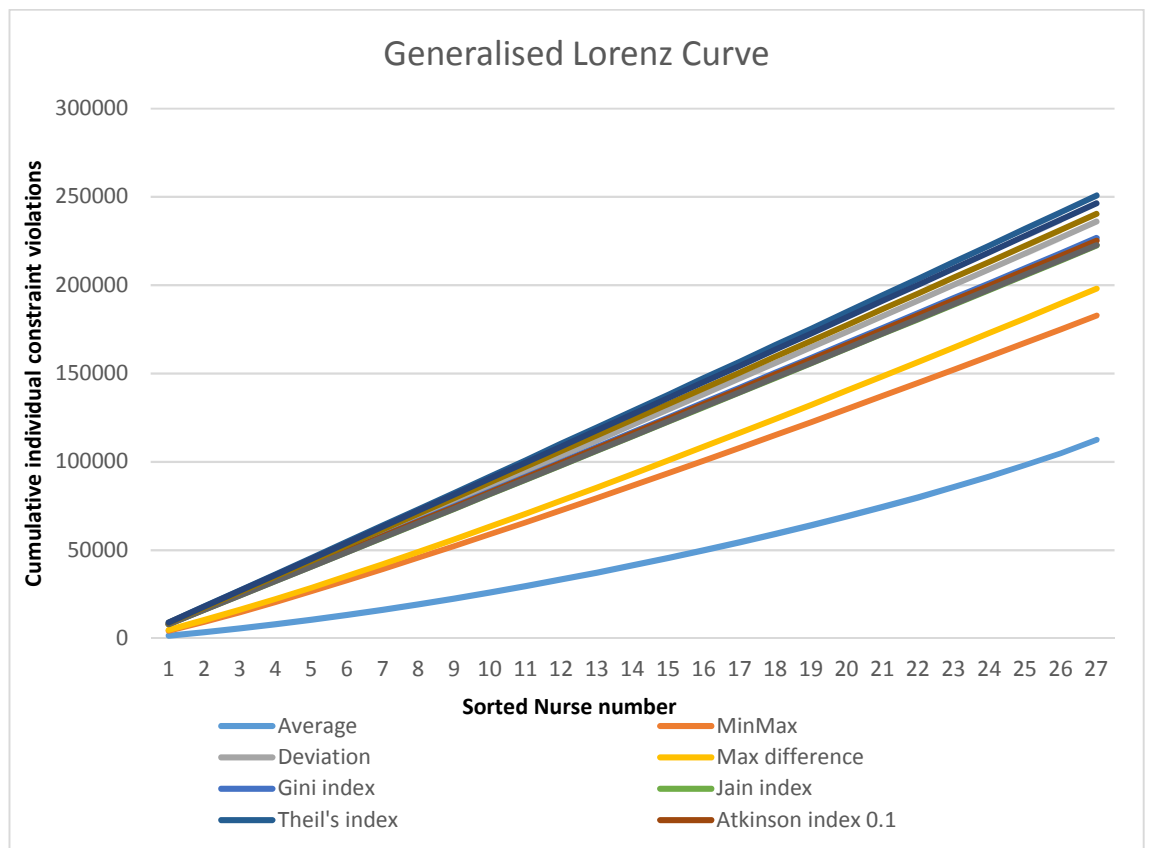

Figure 6-2: Fairness Generalised Lorenz Curve

By observing Lorenz curves, it is apparent that all measures apart from average are not performing well in efficiency. This is normal, since they are only optimizing equality, that is somewhat in contrast with efficiency (Bertsimas et al 2011).

Another aspect of the solution is the computational time. The results in the table below show the time required for the whole run: this includes the calculation of the objective function as well as the calculation of each constraint violation. According to 3.4, it would be expected that Gini index would be more time consuming than the rest of the measures, and simple measures such as average, minmax or Max difference would require considerable less time. However, since the evaluation of the constraint violations is far more time consuming than the evaluation of just the objective function there are no noticeable differences. Also, the moves used in order to generate different solutions (for example swap or delete) effects possible differences in time required for each run. From table 6.2 we can assume that in cases where there is a extensive computational cost until calculating the objective function, computational cost differences between different fairness measures are rather unimportant.

Objective Function	Time required for each run
Average	0.118506236
MinMax	0.11505
Deviation	0.11538
Max difference	0.104949003
Gini index	0.103509555
Jain index	0.100039726
Theil's index	0.09772675
Atkinson index 0.1	0.10222906
Atkinson index 0.9999	0.103170273
Atkinson index 1.00001	0.100070801
Atkinson index 1.9	0.099782517

Table 6-2: Computational cost

6.5 Conclusions

The new measures can produce way more equal results than the ones used in the literature before. In the following Chapters, the three cases using an outranking method are presented. Atkinson index, in its various forms, produced the most equal results.

Chapter 7 Poverty in Nurse Rostering Problem

7.1 Introduction

Nurse rostering problem (NRP), as an application that involves individuals, can utilise the poverty concept.

NRP is modelled as constrain violation problem. Each nurse has some preferences about shifts, working days and a series of other scheduling options. Every one of those options that are not satisfied invokes a constraint violation penalty. The sum of constraint violations for each nurse constitutes the total negative utility imposed on this nurse. The most common evaluation of this model was to sum the total of each nurses constrain violation, as well as the hospital needs for manpower. The sum represents the efficiency of the roster, and for most of the cases it is considered to maximise the happiness within the workforce. However, what it actually does is to try to minimise the nurses' undesirable shifts, thus minimising the total nurses' unhappiness.

While minimising total nurses' unhappiness is a common goal, this goal does still leave cases of extreme unhappiness that are not targeted when trying to minimise the total unhappiness. If someone assumes that there is a point upon which the unhappiness can affect performance where unhappiness bellow that point does not, there is a point on minimising this extreme unhappiness.

The approach proposed in this chapter is slightly different in conception by setting up a threshold on constraint violations upon which a nurses' condition changes from satisfied towards unsatisfied. Depending on the schedule, most nurses are willing to accept certain inconvenient shifts up to a point. This point can be related to the poverty line in economic situations. Thus, a model can be established that resembles poverty, or minimisation of dissatisfaction.

In this context, poverty (or resource starvation) can be used to minimise the extreme (or not so extreme) unhappiness. Like in economics, two are the main issues that arise. The level of the poverty line (or unhappiness line) and how poverty line is utilised to produce some meaningful results. In order to evaluate an allocation, a series of possible measures can be

used. Depending on the desired effect on the various degrees of starved members the appropriate measure should be used.

Every nurses' sum of constrain violation is compared with the poverty line. Since it is a minimisation problem, in contrast with economics, a nurse is considered starved if this sum exceeds the poverty line. This means that the proposed measures should be transformed in order to apply in minimisation problems.

The selection of poverty line can be again a point of discussion. Since there are no hands-on data on what point of nurses' sum of constrain violation a nurse is unable to perform, a series of experiments must be done to discover an acceptable point and investigate the effect on the line on nurses' happiness. Again, poverty line can be absolute or relative. By absolute poverty line it is described as an exact number -arbitrary or not. Relative poverty line is defined as a percentage of a central indicator (average or median) of the current allocation.

Another issue that can arise in Nurse rostering problem, is the evaluation of both the nurses' general happiness and satisfying hospitals' needs. This can be determined by a number of ways. The simplest one is to disregard nurses' or hospitals' needs. Finally, some thresholds can be established for either nurses or hospital satisfaction, and those can be used as hard constraints, Simulating an economic situation, where the amount of resources is limited.

The final approach has some merit in poverty aversion environment. Since the resources distributed in nurses would be accordingly reduced depending on the level of the aforementioned threshold, poverty line definition would be more important.

In this chapter it is presented that poverty concepts can be used in an OR problem. The experiments also show the importance of choosing the correct parameter values.

7.2 Experiments methodology

Since in poverty every individual nurse is compared with poverty line - a fixed value- there is another way to present results. The first of the graphs in every case is a representation of individual nurse constraint violations in the y axis, and the number of nurses (sorted) in the x axis. Presenting results in such a way makes it easier to compare with poverty line. Lorenz curves are included, similarly to other chapters.

The poverty measures that were used were:

Measures Used
Sum
Jain
Headcount
Normalised Poverty Gap
Squared Poverty gap
Watts measure
Income Gap (sen)
Sen measure (with Gini)
Sen Measure (with Jain)
Foster-Greer-Thorbecke 3

Table 7-1: Poverty measures used

The methodology is similar to the previous chapters. Two sets of experiments were performed. The first set of results used only poverty measures, without evaluating any other aspect of the roster. The goal of those tests was to research the use of poverty measures as social welfare measures.

7.3 Results

The poverty line issue was addressed in threefold manner. The first set of experiments were made with a fixed number of weighted constrain violation penalties sum as a poverty line. This approach translates to the real-world analogy that a nurse that has more than a certain number of undesired shifts will be unhappy. Since the constraint violation sum is the depiction of a nurses' unhappiness, it would make sense to limit this unhappiness. The poverty line should be chosen to accommodate those needs. A fixed number could be appropriate in this context. However, since the instances range from simpler with a small number of constrains to more complex with a large number of constrains this approach can lead to a pitfall.

The other set of experiments were performed with poverty line being correlated to how complex an instance is. Poverty line were set to be a percentage of constrain violations from an ideal solution. The results from previous chapters were used in order to find the minimum number of total constrain violations. Different percentages of the best result were used. Even though there is no real-world analogy, those results are based on the tailored needs on each instance. This poverty line approach is in accord to Sen seeing the poverty line set as a percentage of the average income.

Finally, in the third set of results coverage constraints were included. Since a number of assumptions would be needed in order to aggregate individual and coverage constraint

violations, coverage was considered hard constraint: being over an anti-ideal point (see chapter 9) a very large penalty on the objective function was imposed. In this set of experiments, the poverty line was again set at a specific point, similarly to the first set.

All of the following results are from the first instance (emergency).

7.3.1 Absolute poverty

In the first set of results the poverty line was set relatively approximatively. In order to set a poverty line that reflects the needs of nurses based on the constraint violations on certain preferences a lot of research must be done. Firstly, to correctly set the individual preference weights then individual weights to calculate utility for each nurse. Then, evaluation for what is the real weighted constraint violation limit that if surpassed a nurse will be so unhappy that will not be able to function properly.

In order to set a poverty line even with that research, some assumptions are expected to be made. What is the "unhappiness" function, and how it relates to utility? What is this limit that a nurse being so unhappy that will not be able to function properly? Does this limit depend on the individual? In economics, there is a lot of research for setting poverty line, by a lot of scientists. Expecting the same amount of research for a OR problem would be unreasonable

The research done to set the poverty line in the next 2 examples (poverty line 3000 and 5000) was less exhaustive. Since selecting a poverty line based on previous results (see 7.3.2) would be similar to the "relative" poverty approach, poverty line could not be based at previous results, at least not in a direct manner. The next numbers are thus selected in an arbitrary manner. Experiments were done for poverty line set to 1000,3000 and 5000. Results for poverty line set to 1000 were as expected and since no measure managed to have individuals so low they are not being presented in this thesis. Selecting poverty line of 1000-3000-5000 provide a range of poverty lines that allows us to evaluate the performances of measures depending on the setting of poverty line.
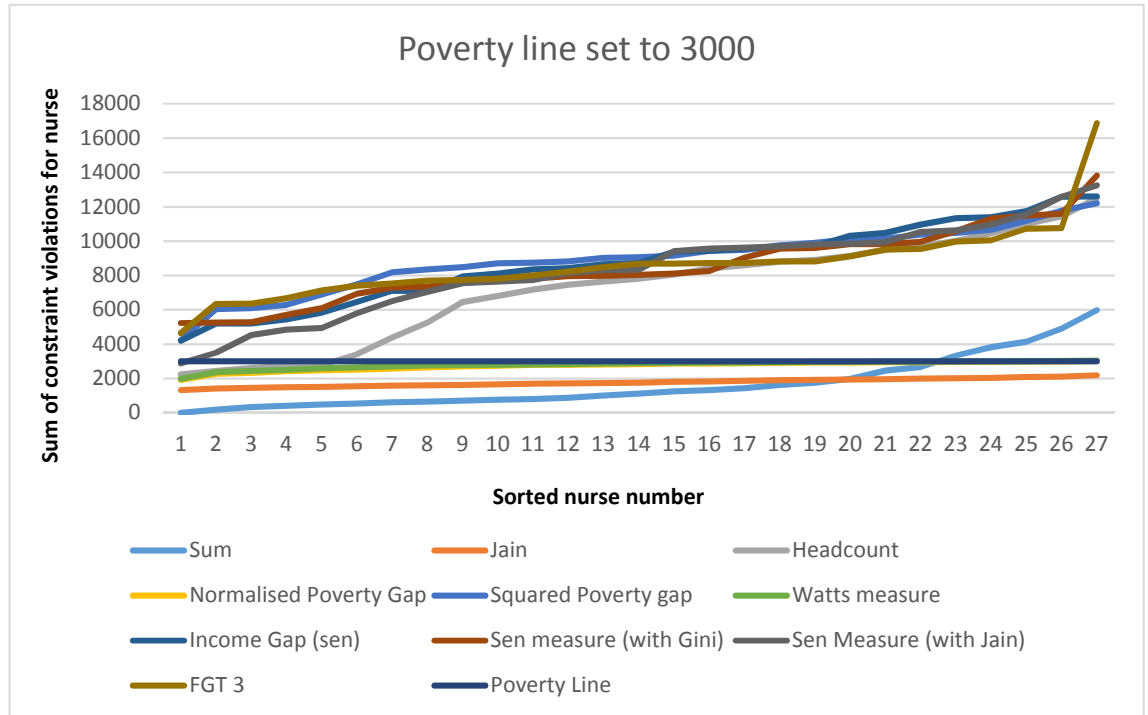
Case 1: absolute poverty, poverty line set at 3000

It is interesting to note that Jain is the only measure constantly below Poverty line.
Normalised poverty gap only has 1 nurse above the poverty line while watts has 3, sum 5.
FGT, income gap, Sen measure with Gini and squared poverty gap are always above poverty
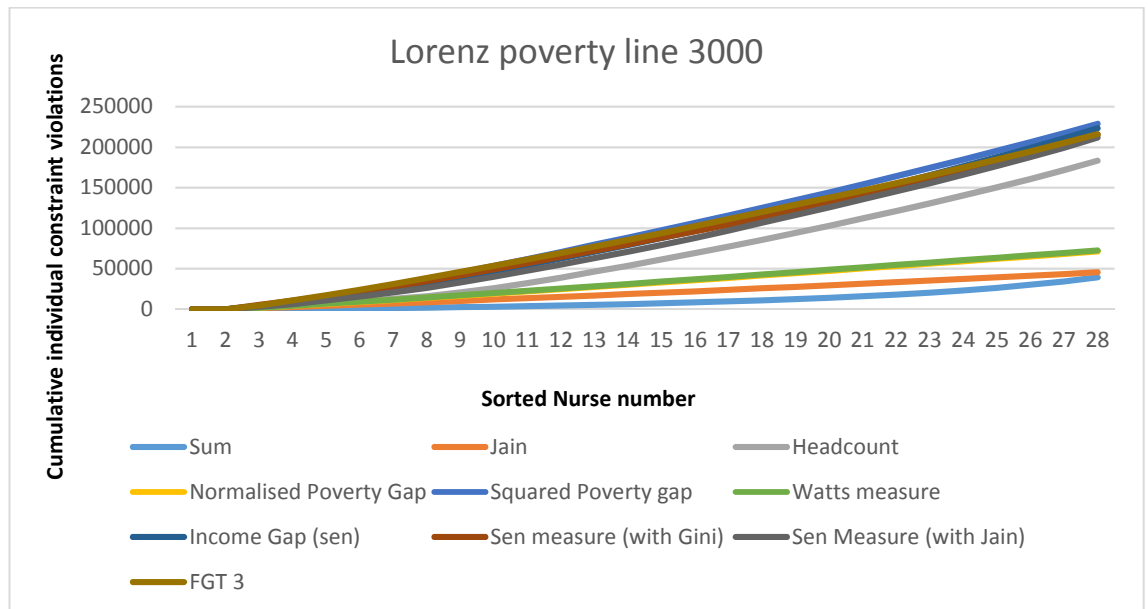line.

The GL curves are quite clear, with not so many points where they cross. Sum is the lowest
curve, therefore the best, followed by Jain, normalised poverty gap, watts measure,

headcount and then Sen measure with Jain. Income gap (Sen), Squared poverty gap, FGT3 and Sen with Gini all cross and are the highest ones.
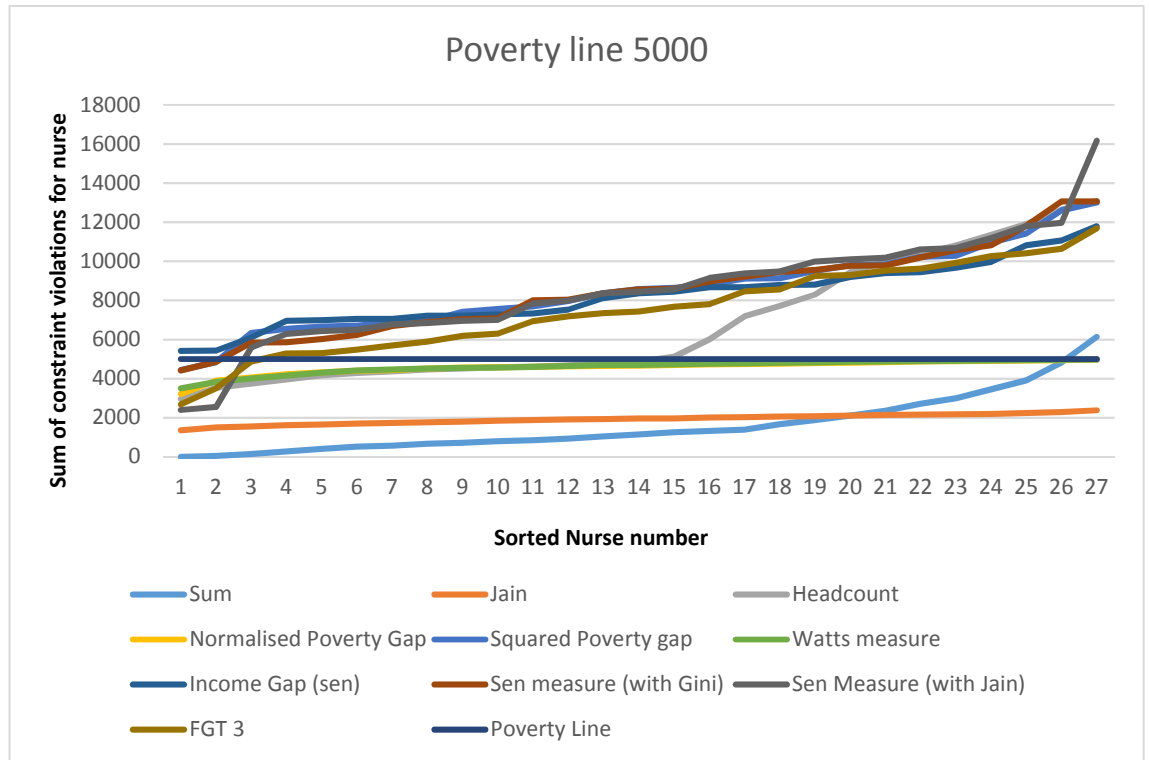

Figure 7-3Poverty results with poverty line set to 5000

With the poverty line set to 5000 more measures have results with many individuals lower than the poverty line. Watts measure, normalised poverty gap and Jain were all below the poverty line. Sum only had one individual above the poverty line. Income gap had all members above poverty line.
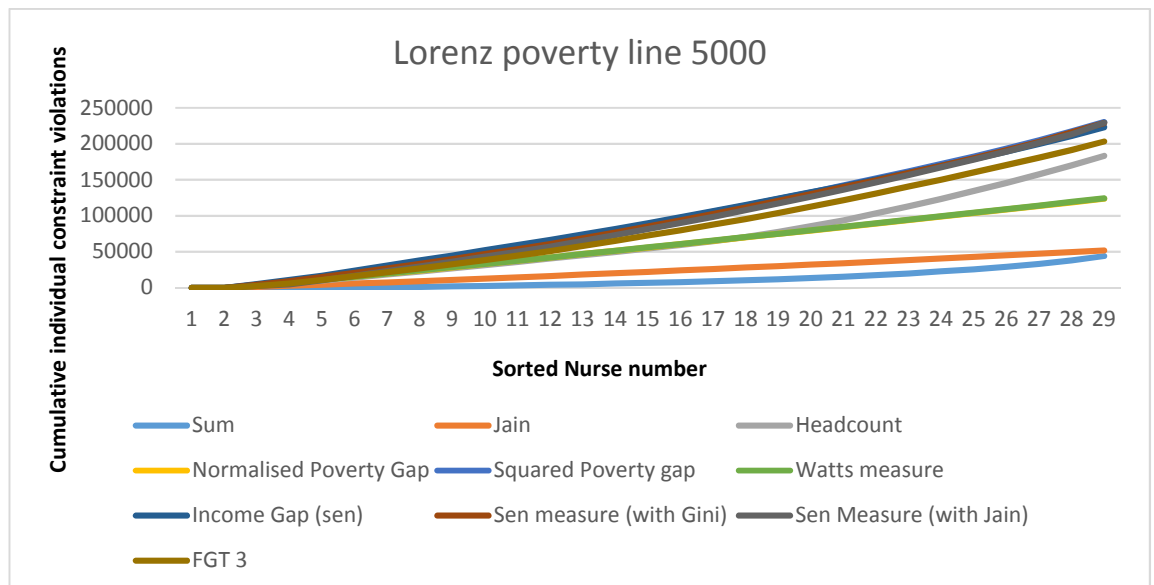
Figure 7-4:Lorenz Curve with Poverty line set to 5000

Sum was the Measure with the lowest Lorenz curve at all points, followed by Jain. Sen measure with Jain, Foster-Greer-Thorbecke 3, Headcount, Normalised poverty gap and Watts measure all crossed. Sen measure with Gini, squared poverty gap and Income gap were the group of 3 worst measures.

Some conclusions that can be drawn from the above experiments are that poverty measures even achieve the goal of being below the poverty line at cases. Utilitarian (sum) and Social welfare function (Jain) (see chapter 8) measures managed to be below the poverty line in more cases than the poverty measures. Measures such as headcount that is unable to take into consideration changes other than the passing of poverty line underperform when drawing the GL curves. A reasoning behind that is that headcount is not continuous, and the search function might hinder the optimisation.

However, it is noted that with higher poverty line more measures seem to reach the goal of being below the poverty line even though that generally does not improve results (as can be seen by the comparison of GL curves). A conclusion that can be drawn is that even though the actual happiness did not increase, setting the levels of unhappiness can affect perceived happiness.

It is also noted that more complicated measures such as Sen measure, squared poverty gap and FGT severely underperform by in cases, not having a single member under poverty line.

7.3.2 Educated Absolute Poverty

The above poverty lines were selected rather arbitrary. There is no clear indication that suggest towards the use of those selected lines. In this section, an educated guess will be made to select poverty lines.

In this section results from previous experiments are being used in order to assign poverty lines. While in relative poverty in economics poverty line is set as a percentage of a central tendency measure of the whole distribution (usually the average) that would not be possible in OR context. In economics the goal of setting a poverty line is to ultimately describe a distribution based on that poverty line, while in this thesis the goal is to decide what distribution is better. Setting the poverty line based on a current distribution, and then evaluating the distribution on the grounds of individual comparisons to the poverty line would create two conflicting goals in the shape of one, and could not direct the objective

function. The expected result would be, depending on the measure, to have some individuals with being way higher than the poverty line and some just below the poverty line.

However, since we have no data on preferences for the threshold of dissatisfaction of nurses' results from previous experiments can be used. Experiments were made (see chapter 9) in order to find ideal and anti-ideal points for several measures, including the average constraint violations of nurses. The anti-ideal point for efficiency was calculated in comparison to coverage constraints. Experiments were made with average nurse constraint violations and coverage constraint violations where weights for each were set high (0.99) and low (0.01) respectively. Those are the best (and worse) acceptable values for average and coverage.

Being consistent with relative poverty theme, the poverty line selected in the next experiments were set to be a percentage of the ideal and anti-ideal point. In the first example (5475) poverty line is the sum of constraint violations of the anti-ideal point divided by the number of individuals, having the poverty line set to be at 100% of the worse possible allocation for nurses. In essence, this poverty line represents the sum of constraint violations each nurse would have in a completely equal distribution of the worse acceptable distribution for nurses' satisfaction.
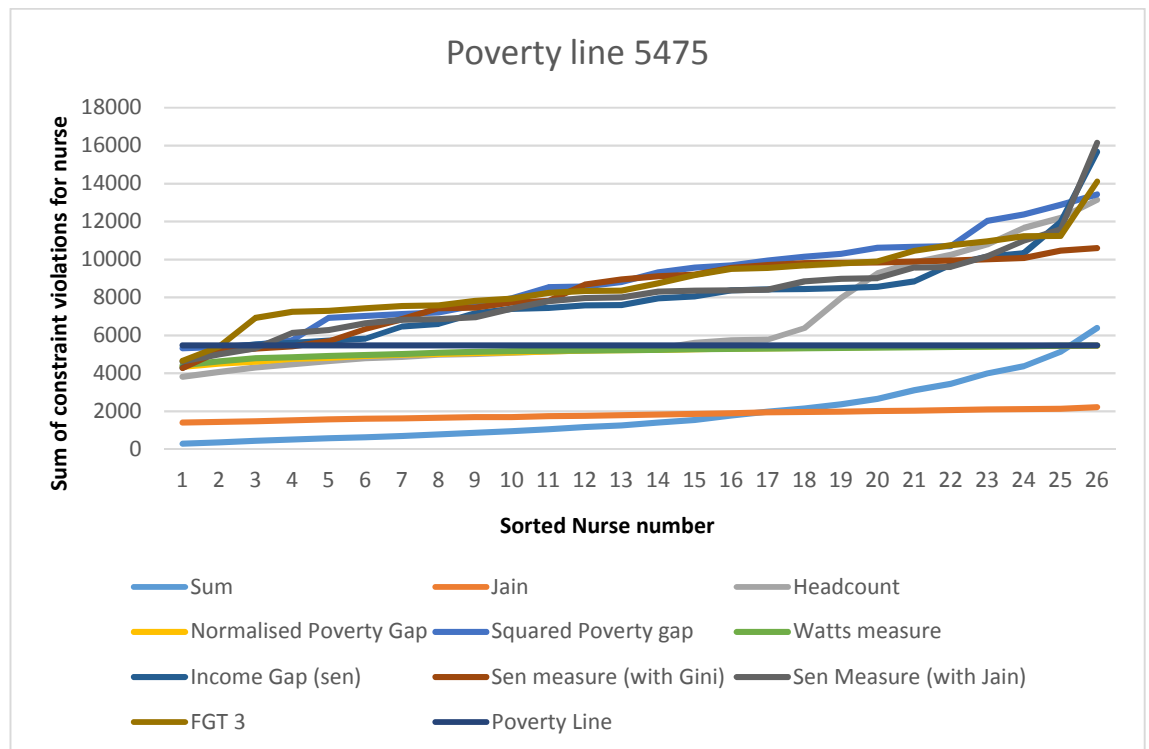

Figure 7-5:Poverty results with poverty line set to 5475

With poverty line set to 5475 Jain, normalised poverty gap and watts measure were the ones having all nurses below poverty line. Sum only had one nurse above the poverty line. All measures had at least 3 individuals below the poverty line.
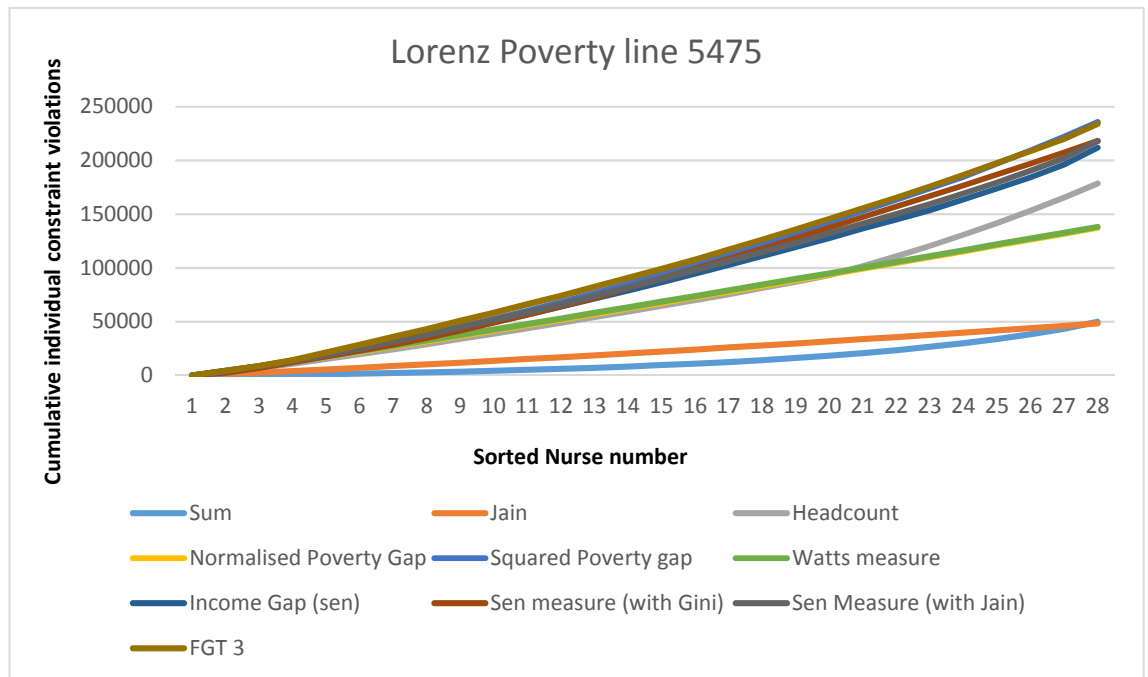
Figure 7-6: Lorenz Curves with poverty line set at 5475.

Sum was the lowest Lorenz curve, followed by Jain. Watts had a high Lorenz curve but by the end it was the 4th lowest one as did Normalised poverty gap being 3rd at the end. Headcount and normalised poverty gap were really close until the 19th nurse where Headcount started to rise.

In the second part of those experiments, the poverty line was set to be the average sum of constraint violations for nurses. This poverty line is based on the best acceptable results for the group of nurses if someone only focuses on the efficiency of their roster. In essence the poverty line is the sum of weighted constraint violations that each nurse would have in a completely equal best acceptable distribution. It would be expected that this poverty line is relatively low, since it is based on the best results for nurses.
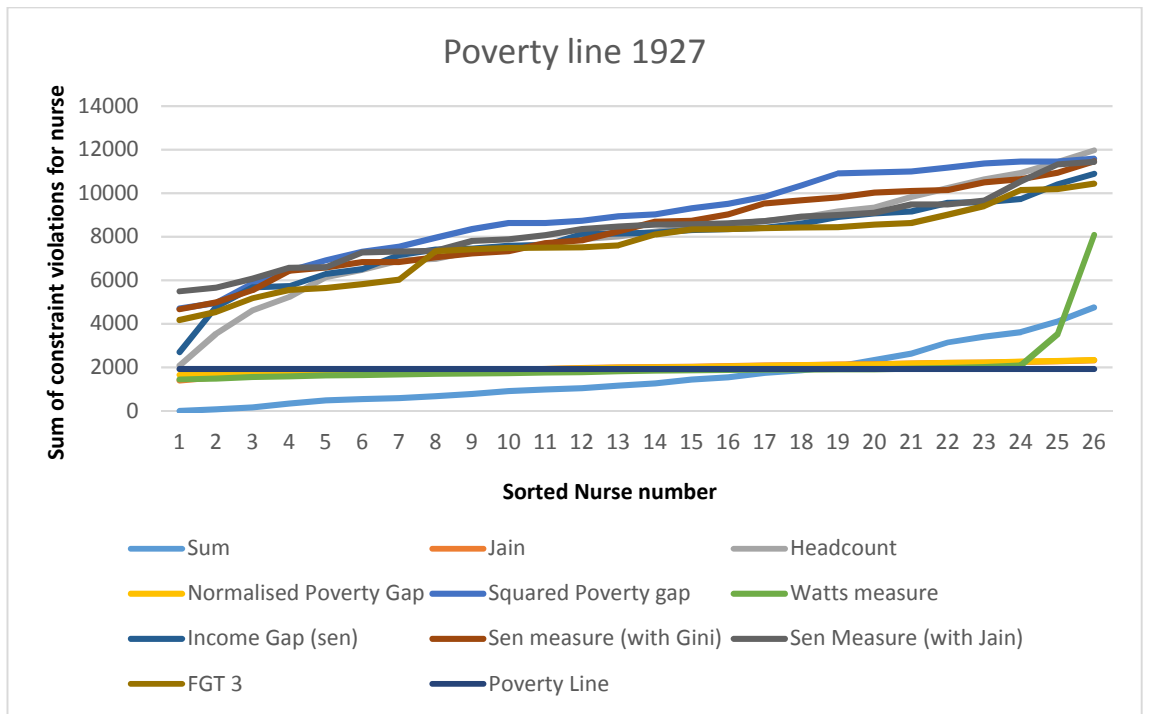
Figure 7-7:Poverty results with poverty line set to1927

Since 1927 is a relatively low for this instance, it was expected that not many individual nurses were above poverty line. Not even Jain and Sum that were entirely below poverty line in the other experiments were, with sum having nine individuals above poverty line, Watts seven and Jain 17. Headcount, Income gap (Sen), FGT3, Squared poverty gap and Sen measures (with Gini and Jain) had no individual above poverty line.
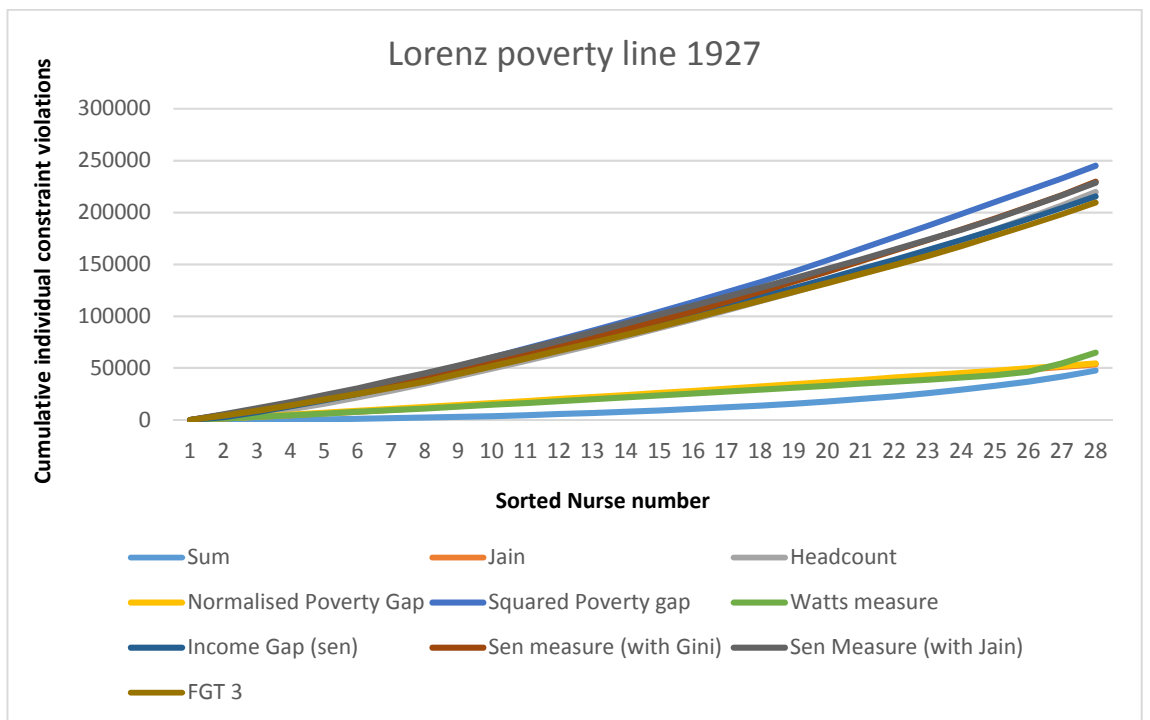
Figure 7-8: Lorenz Curves with poverty line set at 1927.

Sum was the Lowest curve when setting poverty line at 1927, followed by Jain, Watts and normalised poverty gap. Headcount (that performed better at first nurses) income gap (better at the first and last nurses) and FGT followed crossing between them. Sen measure with Gini, Jain and Squared poverty gap followed being higher than all curves.

The results are similar to the previous section. Having poverty line set higher allows for more individuals to be below poverty line, but did not really improve Lorenz Curves. Some Lorenz curves are better when setting higher poverty line than lower. This might be happening because of the search function: if there is not an apparent improvement in the objective function leading the direction to a better solution, the search stagnates and ultimately produces worse results. Better performing poverty measures such as Watts and Normalised Poverty gap converge to the poverty line. Again, the perceived improvement on the extremely unhappy does not mean an improvement on the actual constraint violations.

7.3.3 Experiments with resource constraints

The second part of the experiments is made to create a roster that is effective to both the nurses and the wards. The procedure is similar to the previous chapter. Poverty line were selected to be 3200. This number is somewhat in the middle between the highest and lowest poverty lines used in this thesis. The poverty models were tested in contrast to the ward efficiency. In order to simplify things, Ward coverage was considered a hard constraint, set at a previous found anti-ideal point.

The difference between those experiments with the previous chapters is the inclusion of coverage constraints. In the previous chapters coverage constraints were ignored. However, the approach of coverage constraints in this chapter is different than the others.

In chapter 8, coverage constraints were aggregated with nurse constraint violations. Even though they were treated as different, the sum of weighted coverage and nurse constraint violations were used as the objective function. In chapter 9, coverage was a separate goal to optimise, and then depending on the model (minmax or weighted) was included in the objective function.

The concept of coverage constraint violations in this chapter is entirely different. Coverage cannot be treated as another nurse, setting poverty line to be evaluated upon. That would create difficulties in creating the model, such as the relative weight of nurse and coverage constraint violations, the poverty line of coverage etc. so coverage constraints were treated different than other chapters. A limit upon that once surpassed a great penalty on the

objective function was imposed, making it a hard constraint. However, if this limit was not surpassed coverage was not included in the objective function.
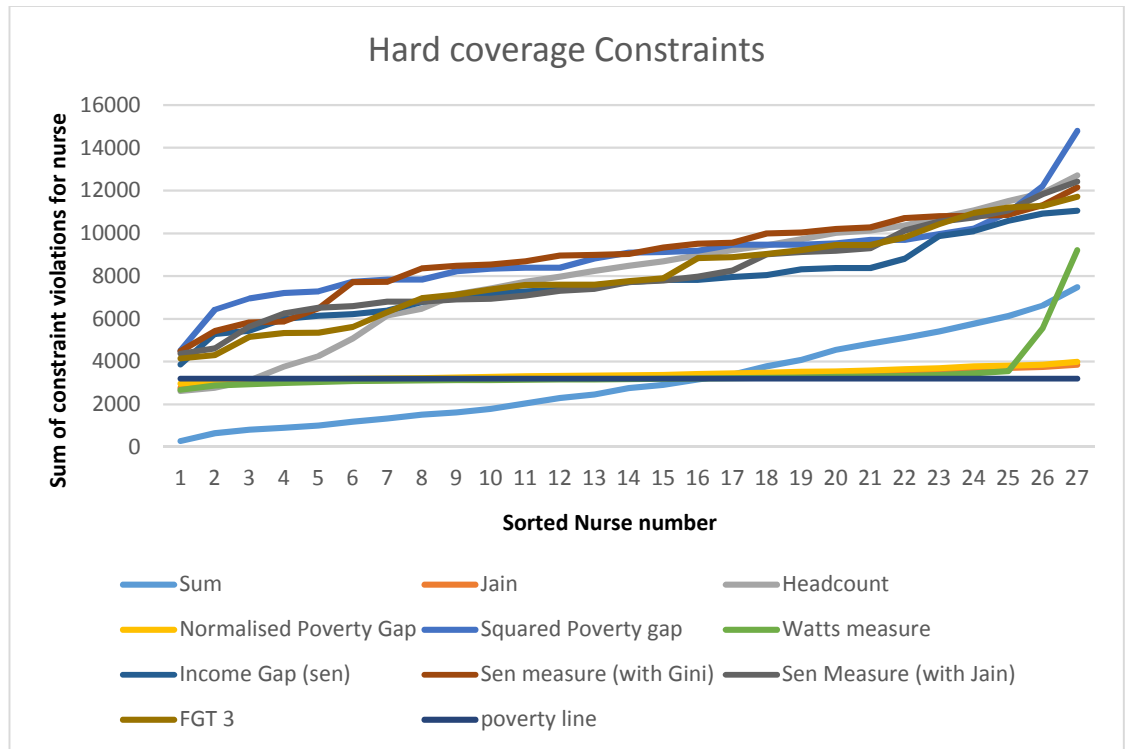
Figure 7-9: Poverty results with Hard constraints

Having low and hard coverage constraints even less individuals were above the poverty line. 11 nurses from sum, 12 from Watts, and 16 nurses from Jain measure were above the poverty line. Measures such as Income Gap (sen), FGT3, Sen measure (with Jain and Gini) and squared poverty gap had no individual below the poverty line.
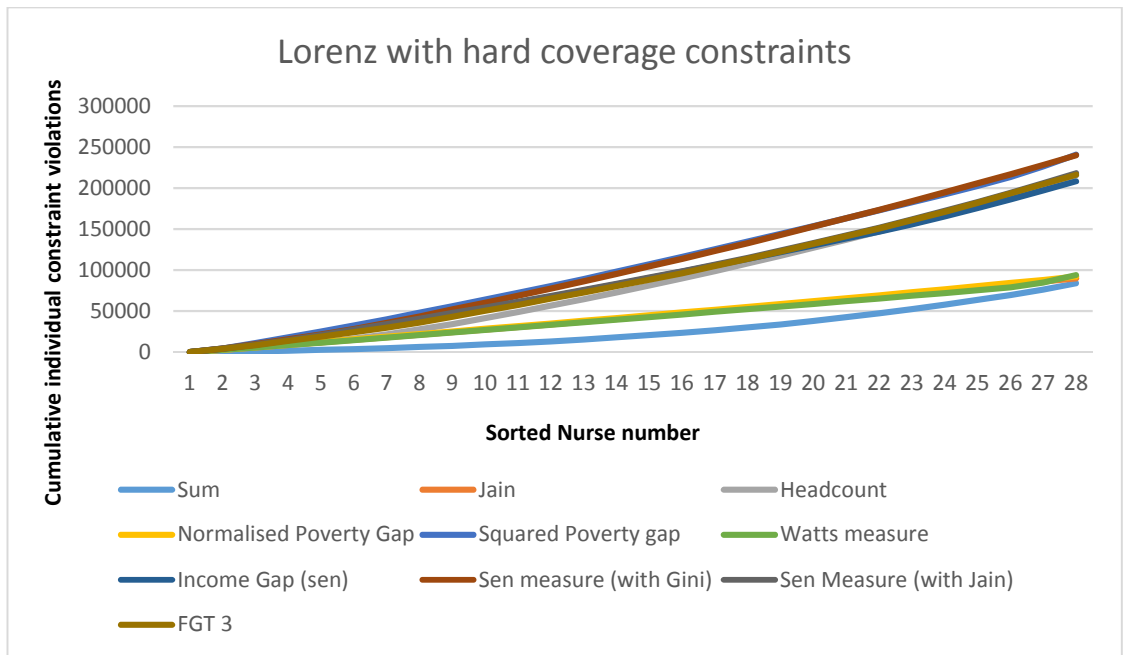
Figure 7-10:Lorenz Curve with Hard Coverage Constraint

As in previous cases, Sum was the one with the lowest Lorenz curve at all points. Watts, normalised poverty gap and Jain form another group, only surpassed by headcount in the first 3 nurses. Sen measure with Jain and Gini, as well as squared poverty gap were the higher Lorenz curves.

The results on poverty under hard coverage constraints confirm the suspicions: overall results are slightly worse than the cases where no hard constraints on coverage were implemented. However, in one case a feasible solution was not found (squared poverty gap).

Generally, imposing hard coverage constraints did affect the quality of nurse results, in both minimising unhappiness and the results presented in GL curves. This could go well with a real world analogy: having limited recourses makes poverty harder to deal with than having infinite recourses.

Measure	Coverage Constraints
Sum	8000
Jain	8000
Headcount	7320
Normalised Poverty Gap	7940
Squared Poverty gap	86000
Watts measure	8000
Income Gap (sen)	1800
Sen measure (with Gini)	1800
Sen Measure (with Jain)	2000
FGT 3	1800

Table 7-2: Coverage constraints violations

7.5 Conclusions

Most of the poverty measure results were sub-par, producing worse results than fairness and the efficiency measure. However, a simple poverty measure, poverty gap, performed relatively well in every instance. The nature of poverty measures, being non-convex is related to the way our algorithm works.

General conclusions that can be drawn from this chapter is that setting the poverty line affects the perceived unhappiness but might not affect the actual results. Simpler measures such as poverty gap and up to a point headcount performed better than the more complicated ones, and the search function might have had a role on that.

Generally, poverty alone cannot be used in order to produce efficient results, and might be better to use central tendency measures in order to produce poverty-aversion distributions. However, poverty can still be used as a part of an objective function with other goals.

Chapter 8 Social Welfare Functions

8.1 Introduction

Social welfare function, a term introduced by Bergson (1938) corresponds to a complete evaluation of a society. Seminal work from Sen (1973) and Atkinson (1970) set the foundation on this work.

As shown, having perfect fairness does not guarantee good solutions. Fairness is about the distribution of resources, not the amount of the resources distributed. The results from chapter 6 agree with this Hypothesis. Even though the fairness measures produced fair rosters, those rosters are not efficient, and could not be applied in practice.

For example, in the results discussed in 6.4, even though Atkinson measures performed really good in fairness, they all underperformed in every individual compared to even the worst-off individual in Average. All nurses' sum of constraint violations was higher than the sum of constraint violations of the nurse that had the most in the allocation produced from Average measure. It can be assumed that even with a roster being fairer, a community of nurses would not accept a roster that was worse in every individual comparison.

Kaplow and Shavell (2003) also discussed about fairness not necessary linking to Pareto optimality. In fact, there are a lot of cases that fairness is correlated with reducing the happiness of the individuals involved (Kaplow and Shavell 1999). They advocate the use of Social Welfare Functions (SWF) for decision making.

In order to address this issue fairness measures were combined with an efficiency measure, namely sum or average (Sen 1971, Dagum 1990, Atkinson 1970). Some arguments for Welfare instead of fairness can be given in Kaplow and Shavel (2000).

In SWF the evaluation of a society is made upon a function of utility. The difference between SFW and inequality aversion is that inequality only studies the differences between utilities of individuals, while SWF also examines the level of individual Utility.

According to Rawls (1971) an evaluation of an allocation can be written as a function of the individual utility in an allocation.

$$w = w(x_1, x_2, \ldots x_n) \ (1)$$

This formula indicates that the welfare w of a society is a function of utilities x_i of individuals in a society.

According to Sen, welfare is also an expression of total utility and the distribution between individuals, such as:

$w = w(\mu, I)(2)$. where w is welfare, μ is the mean utility, and I is the inequality between individuals. Multiplying (1) with mean utility and dividing all individual incomes by mean utility it would lead us to this expression

$w = \mu w(\frac{x_1}{\mu}, \frac{x_2}{\mu}, \dots \frac{x_n}{\mu})$ (3)

In case of perfect equality between x_i the welfare of this distribution would be equal to the mean utility.

And in turn to $w = \mu(1 - I)$ (4)

Where I is an inequality index bounded between 0,1. This inequality index traditionally is Gini index. However, there is no reason not to select another bounded inequality measure.

Since I is limited between 0 and 1, the welfare function considers an allocation to be as good as the sum of the individual utilities in case of perfect equality. However, in case of perfect equality and low average utility, the welfare of a society would be considered at best average, as we can see from chapter 6.3.

In this research, we expand the definition of 3, and more bounded Fairness measures are used instead of just Gini index. Atkinson (1970), Jains (1984) and Theils (1967) measure are used.

8.2 Models

As in fairness, there are properties that SWF should satisfy. Some of them are identical (transfer principle, scare invariance) however the difference is that SWF should target pareto optimality.

Social Welfare	A-fairness	SW with Theil	SW with Gini	SW with Jain
Transfer principle	✓	✓	✓	✓
Scale invariance	✓	✓	✓	✓
Principle of population	✓	✓	✓	✓
Decomposability	✓	✓	✗	✓
Pareto	✓	✓	✓	✓
Boundedness	n/a	n/a	n/a	n/a
Complete ordering	✓	✓	✓	✓
Landau notation	O(n)	O(n)	O(nlogn)	O(n)
Alg analysis##	10n+3a*n	20n	$12n + 8n\log_2 n - 4 * 2^{\log_2 n}$	10n

Table 8-1: Social welfare measures

Some observations about the properties of the above measures. All of them satisfy the transfer principle, scale invariance, principle of population, can provide a complete ordering, and strive for pareto optimality. None of them is bounded: the reasoning is that welfare can have no upper limit, subject to constraints. Gini is not decomposable, transferring this property from Gini measure, that is not always decomposable. The computational cost is the same as with fairness properties, since the mean can be calculated in one go, while calculating the fairness measures themselves.

The models used were following. The first five measures were introduced in Martin et al (2013).

Measure	Objective function
Average	$$WoOld = \sum_{i=1}^{i=N} y_i$$
MinMax	$$FoaOld = maxy_i + \sum_{i=1}^{i=N} y_i$$
Deviation	$$FobOld = \sum_{i=1}^{i=N} |y_i - \mu| + \sum_{i=1}^{i=N} y_i$$
Max difference	$$FocOld = (maxy_i - miny_i + \mu) * N$$
Sum of Squares	$$FodOld = \sum_{i=1}^{i=N} (y_i)^2$$
Gini index	$$Foa = \sum_{i=1}^{N} y_i * \left\{ \frac{2n+1}{n} - \frac{2\sum_{i=1}^{N} i*y_i}{n\sum_{i=1}^{N} y_i} \right\}$$
Jain index	$$Wo = \frac{\sum_{i=1}^{N}(y_i)^2}{\sum_{i=1}^{N} y_i}$$
Theil's index	$$Fob = \frac{1}{N * \log(N)} \sum_{i=1}^{i=N} y_i * \log\left(\frac{y_i}{\mu}\right)$$
Atkinson index 0.1	$$Foc = \left(\sum_{i=1}^{i=N} \left(\frac{\mu^2}{y_i}\right)^{(0.9)} \right)^{\left(\frac{1}{0.9}\right)}$$
Atkinson index 0.9999	$$Fod = \left(\sum_{i=1}^{i=N} \left(\frac{\mu^2}{y_i}\right)^{(0.000001)} \right)^{\left(\frac{1}{0.000001}\right)}$$
Atkinson index 1.00001	$$Foe = \left(\sum_{i=1}^{i=N} \left(\frac{\mu^2}{y_{i_i}}\right)^{(-0.000001)} \right)^{\left(\frac{1}{-0.000001}\right)}$$
Atkinson index 1.9	$$Fof = \left(\sum_{i=1}^{i=N} \left(\frac{\mu^2}{y_i}\right)^{(-0.9)} \right)^{\left(\frac{1}{-0.9}\right)}$$

Table 8-2: Social welfare measures

The fairness measures were adjusted in order to be completely fair when then measures are 1. In order to construct the Social welfare function, they were multiplied by the average. This way the measures are maximised in complete fairness, and when the average is maximised.

While something relevant has been done with Gini index (eg Atkinson 1970) Jain and Theil have not been included in such a way in SWF.

An objective that someone could raise for those measures is measure 2 and four do not evaluate the whole allocation, but rather the outliners. Measures 6 to 8 were adapted from fairness measures using (3) to include effectiveness. Measures 9 to 13 are the ones proposed by Atkinson (1970).

8.3 Methodology

In order to create a roster acceptable by both nurses and hospital some assumptions must be made. To begin with, the lengths of hospitals flexibility in wards coverage is a matter of discussion. Some employers are willing to make small compromises in order to keep employees happy and productive. Another possible way to approach this issue, is the negotiating power of the nurses, and the enforcement power of the hospital's management. Both of those approaches imply tradeoffs between two goals.

The objective function used for those experiments was

$$OF = w_n U(NC) + w_{cov} U(CC) \text{ (1)}$$

Where w_n the weight for nurses' welfare satisfaction, $U(NC)$ the normalised function of nurses' welfare, w_{cov} the weight for hospitals needs coverage and $U(CC)$ the normalised hospital ward coverage. The sum weights of nurse and ward coverage satisfactions equal to 1.

$$w_n + w_c = 1$$

Different values of w_c and w_n were chosen, similar to Jones (2011) (Table 8.3). The lower weight (0.05) was selected in order to heavily focus on the nurses' satisfaction. The higher value of ward coverage weight reflects the power of hospitals upon their employers. The intermediate values (0.4, 0,6) to show compromises between nurses' satisfaction and hospitals' needs.

Coverage (w_c)	SWF (w_n)
0.05	0.95
0.4	0.6
0.6	0.4
0.95	0.05

Table 8-3: Weight Grid

The values of $U(NC)$ and $U(CC)$O Range from a few thousands to a couple of hundred thousand. That values are the weighted sum of constraint violations, and the score of Hospital satisfaction (coverage constraints) and nurse satisfaction (nurses' constraints). In order to be able to weight coverage and SWF results a normalisation must be performed. The reasoning behind this is that if there is no normalisation the values of the measure that ranges higher will have a greater impact on the overall objective function (1) over the ones that range lower.

Tests were run to find ideal and anti-ideal points for both SWF and Coverage. To find the ideal points for SWF runs were performed with coverage weights set to 0.9999 and 0.0001. The same procedure was followed to find the ideal and anti-ideal points in coverage. Coverage and SWF values were tailored according to the following formulas.

$$U(CC) = \frac{cov - cov_{best}}{cov_{worst} - cov_{best}}$$

$$U(NC) = \frac{(SWF - minSWF)}{maxSWF - minSWF}$$

$U(CC)$ is the value of hospital's coverage satisfaction. cov is the value of the sum of the weighted constraint violations for coverage in this particular roster. cov_{best} is the ideal point of coverage constraint violations that was found using weight as 0.9999. cov_{worst} is the anti-ideal point of coverage constraint violations obtained by using coverage weight as 0.0001. cov_{worse} is larger than cov_{best} since it Is a minimisation problem.

Similar procedure was made to normalise nurse's satisfaction ($U(NC)$. SWF was the sum of weighted constraint violations for this allocation. $minSWF$ is the minimum constraint violations for this specific measure, obtained by using nurses satisfaction weight as 0.0001 and coverage weight as 0.9999. $maxSWF$ is the maximum constraint violations for this specific measure, obtained by using nurses satisfaction weight as 0.9999 and nurses satisfaction weight as 0.0001.

8.4 Evaluation

In order to evaluate both fairness and effectiveness, Lorenz and generalised Lorenz curves were used, described in chapter 4. However, coverage cannot be included in Lorenz or generalised Lorenz curves, since more assumptions are needed on where the coverage importance lies.

Two different set of results are presented. In the first set different measures are compared using the same weights in coverage and nurses' satisfaction weights. Those comparisons' goals are to examine the relative performance between the measures, and potential differences between them when the weights of nurses and coverage change.

In the second set comparisons between results produced by the 3 SWF functions of the form $w = \mu(1 - I)$ using different weights. This is to analyse the sensitivity of each measure using different weights of coverage and nurse constraint violations.

8.4.1 Lorenz Curves

Lorenz curves are again used in order to compare the fairness aspect of the solutions. The lines that are closer to equality line are considered to be more equal – the ones that are drawn upper. All The results presented are from the first instance.
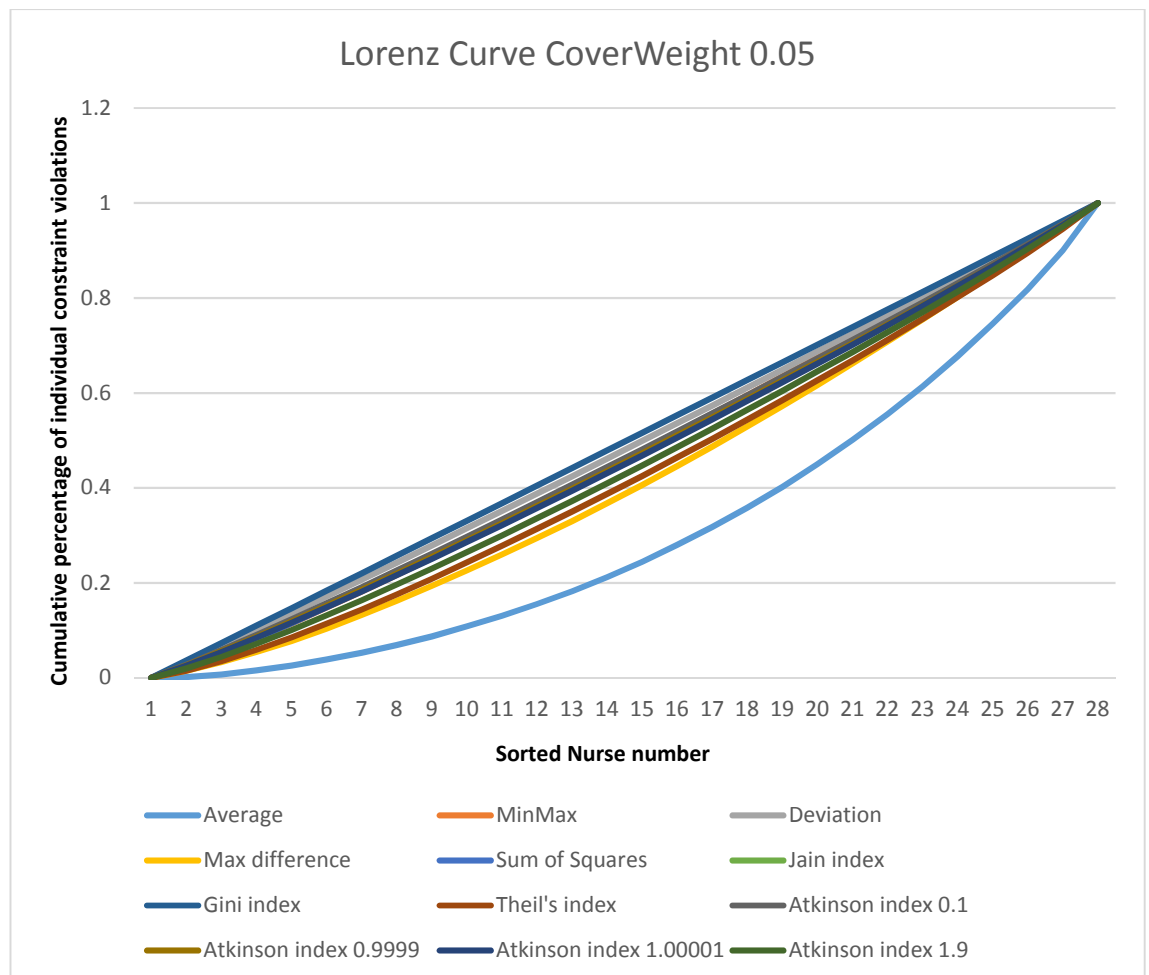

Figure 8-1: Lorenz Curve CoverWeight 0.05

When setting the weights of nurse satisfaction to 0.95 and coverage to 0.05, the most equal results according to Lorenz curves are the ones produced by Gini index, without crossing with

any other line. The second best are the ones produced by Deviation, again without crossing with any other line. The next lines are crossing: Atkinson index 0.1, 0.9999, Jain index, Atkinson 1.00001, Square sum and MinMax by order of performance in the worst of members. Average is the one lower, followed by Max difference and then Theil, without those lines being crossed.

Since the value of weights for nurses' satisfaction is the highest in this example, results are more consistent. For example, someone could expect that average would not perform well in fairness.
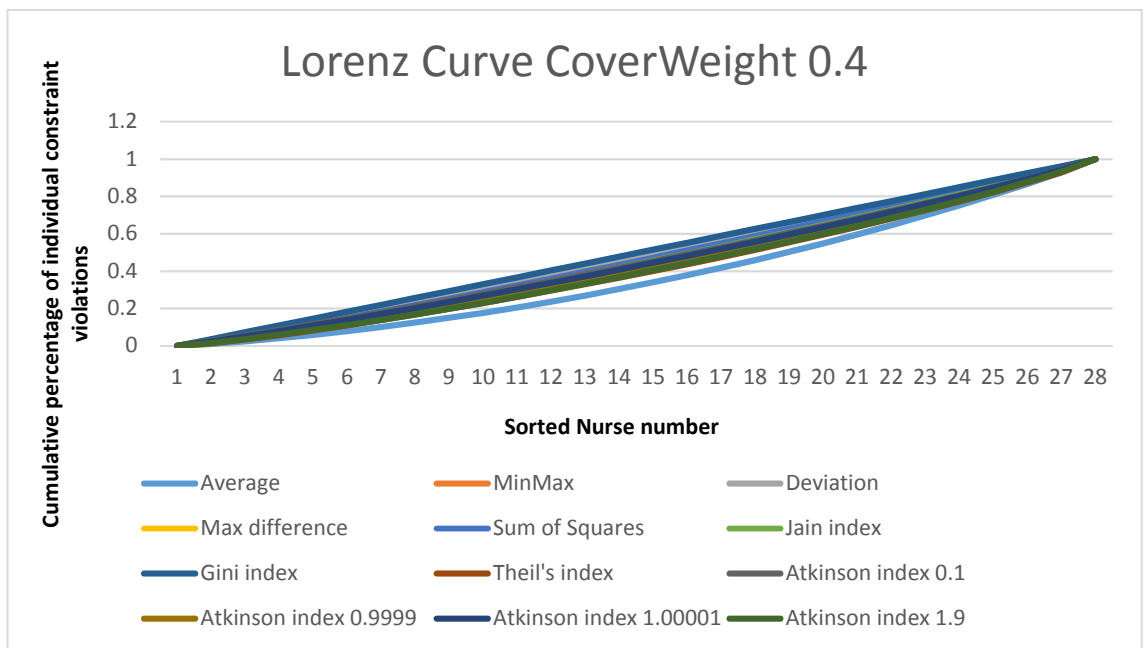
Figure 8-2: Lorenz Curve CoverWeight 0.4

Gini index is again the higher curve, without crossing with any other line. The other curves are crossing in various points. The only safe outcome that could come out of this example is that Gini is producing the fairest results, and the others are incomparable.
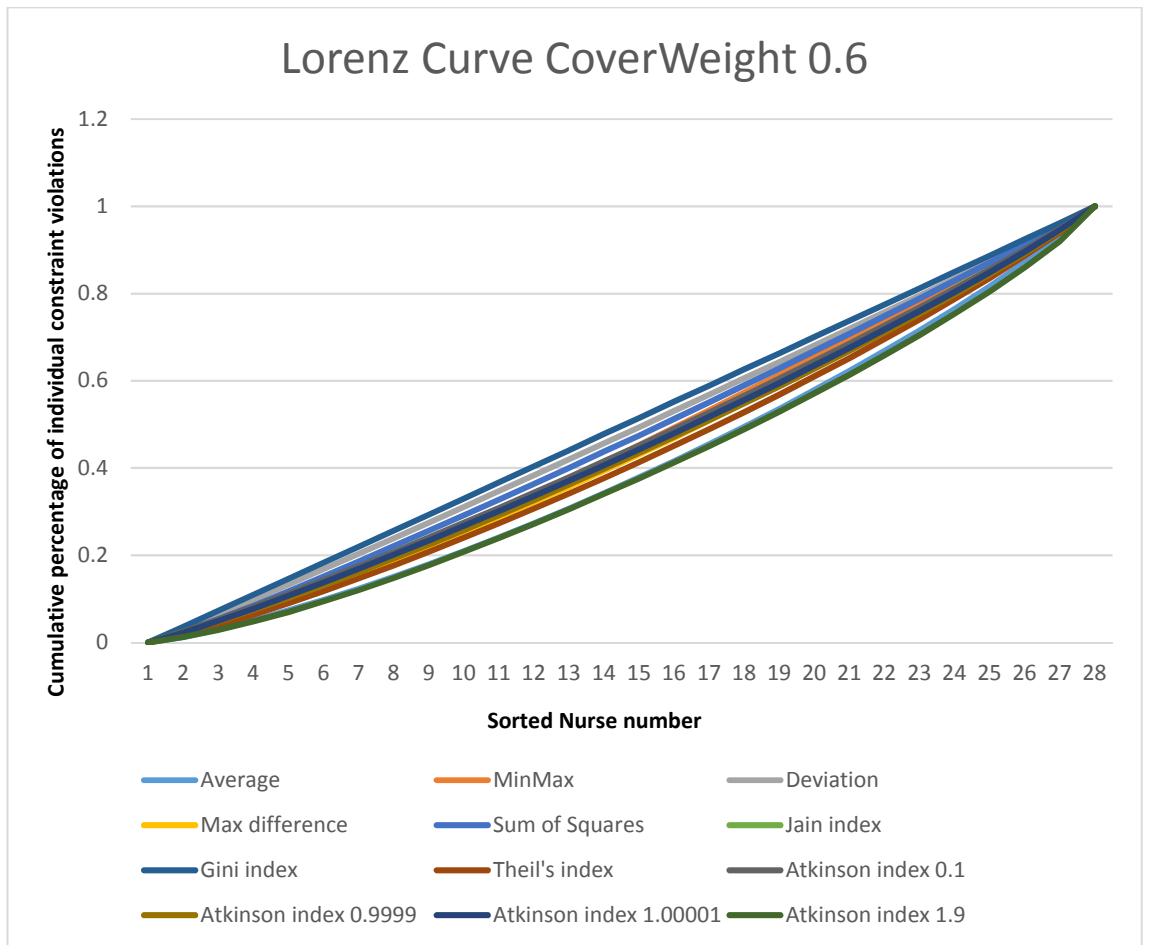
Figure 8-3: Lorenz Curve CoverWeight 0.6

Gini index again produces the fairest results, being the higher curve without crossing with any other curve. The rest of the curves cross. The ones performing worse are Atkinson index and Average (even though their lines cross, their lines do not cross with any other line). Since the weight of nurses' satisfaction starts to be lower than hospital satisfaction irregularities start to happen.
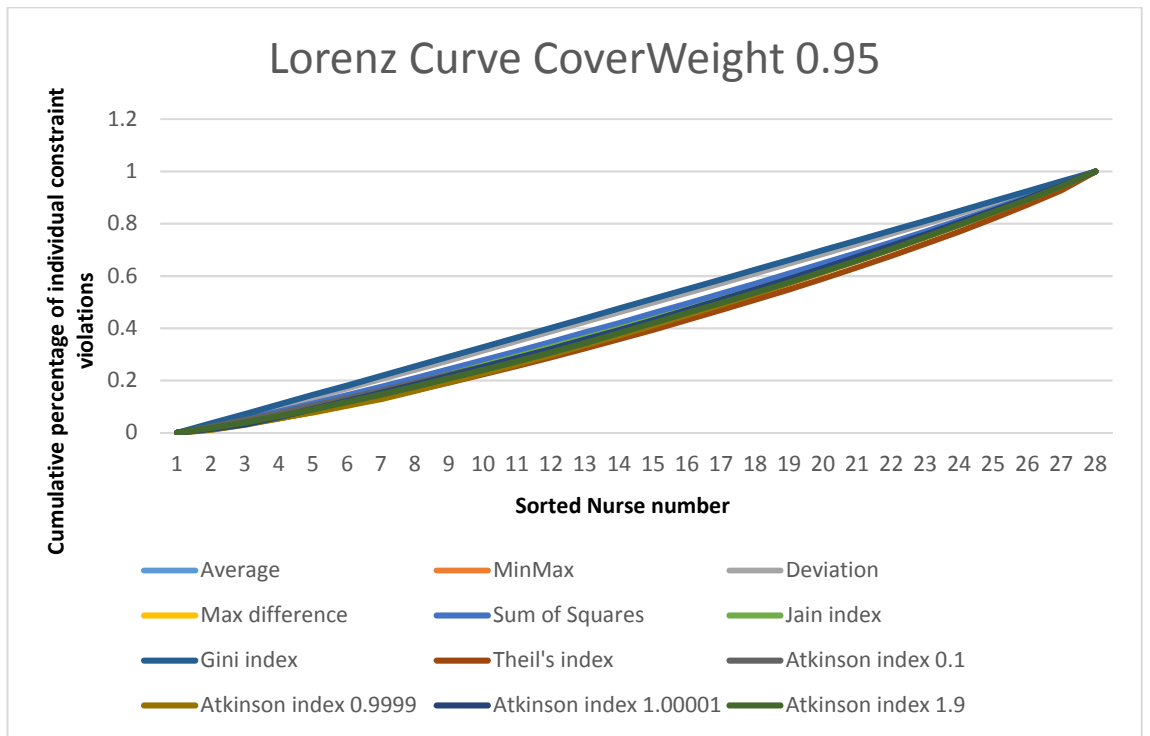
Figure 8-4: Lorenz Curve CoverWeight 0.95

In the final experiment, where nurses' satisfaction starts to be negligible compared to coverage constraints Gini again provides with the fairest results, followed by Deviation. However, the rest of the results become incomparable: At some point average outperforms in the aspect of fairness all but four measures. That's because coverage constraints are so much more important than welfare results that the small differences between measures are not relevant.

8.4.2 Generalised Lorenz curves

Generalised Lorenz curves were used to compare results to compare the "welfare" aspect of the measures. The lines that are lower are considered better, since they represent the cumulative sum of weighted constraint violations.
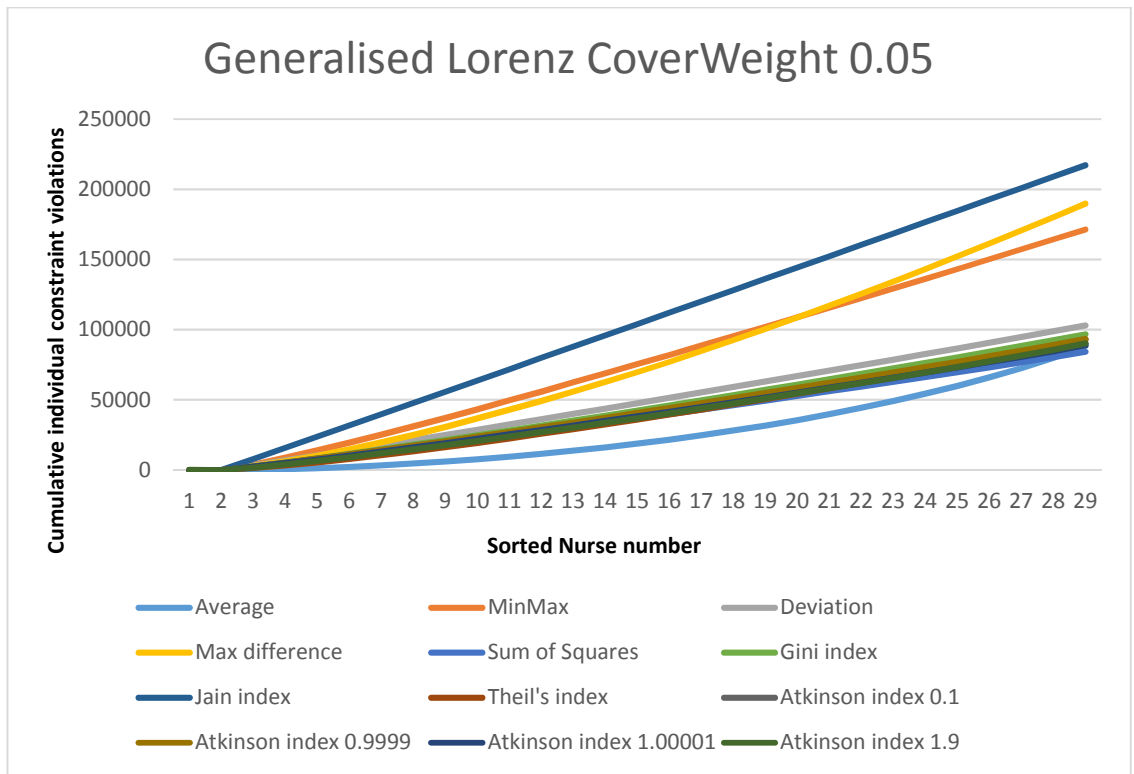
Figure 8-5: *Generalised Lorenz CoverWeight 0.05*

In the first case, where the nurse's satisfaction is most important most of the results are incomparable and their Generalised Lorenz (GL) curves were crossing. The only exception is that Gini index is by far the worst, not being crossed by any line. The average outperforms all other measures for the majority of cumulative nurses, however is crossing with one (square sum). The results from Theil's index, Sum of squares, and all Atkinson measures are crossing, and in the end not far from average. Jain index, max difference, Deviation and min max are clearly the worse (excluding Gini).

Since nurse's satisfaction is much more significant than ward coverage some conclusions can be drawn. Welfare performance is antithetical with fairness. Measures such as average that performed bad in fairness are performing good at welfare and vice versa. There are some clear groups of performance based on welfare: even though the GL curves are crossing, they do not intersect so much as the next cases.
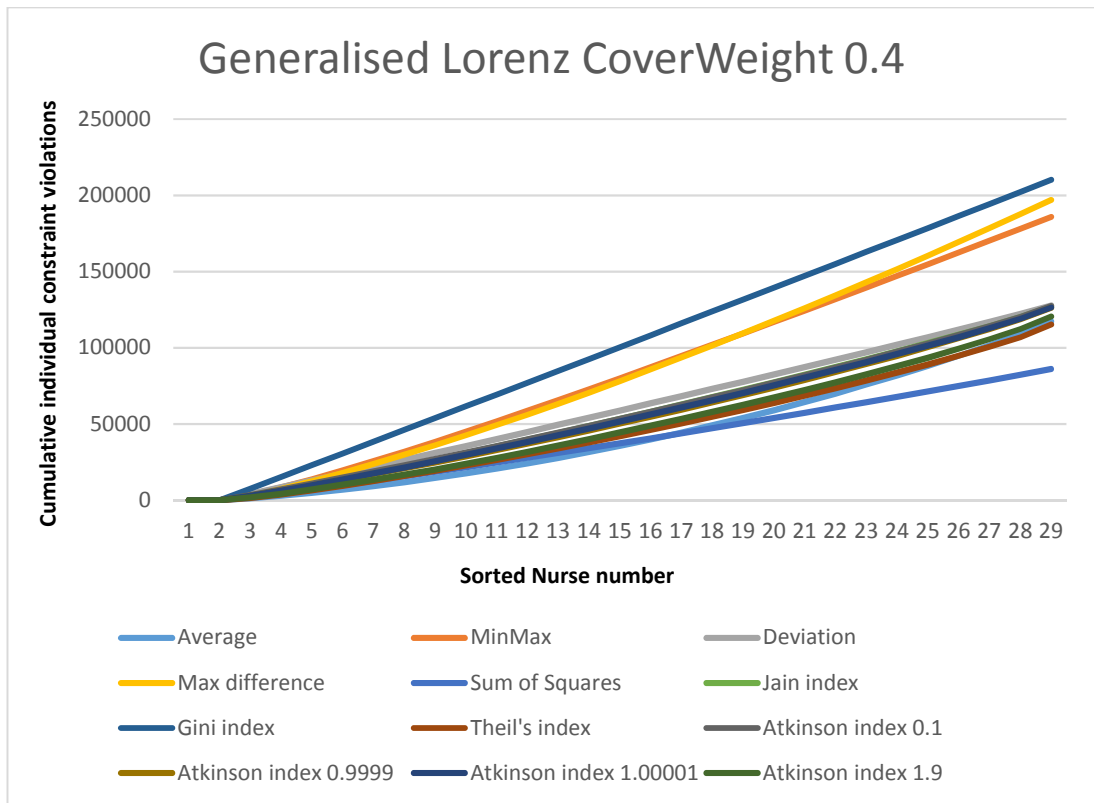
Figure 8-6: Generalised Lorenz CoverWeight 0.4

In this case the results from Theil, average, Sum of squares and Atkinson 1.9 cross and are close for the most part. Gini is the only one that can its GL curve do not intersect with any other, going higher than any other. Minmax and max difference are the second higher group, even though max difference intersect at the first nurses and even outperforms other measures (coming 4rth best in the first nurse).

As the weight for nurses' happiness star to lower, the results become more inconsistent. However Theil, Average, Square sum and Atkinson 1.9 still have high performance as in the previous example.
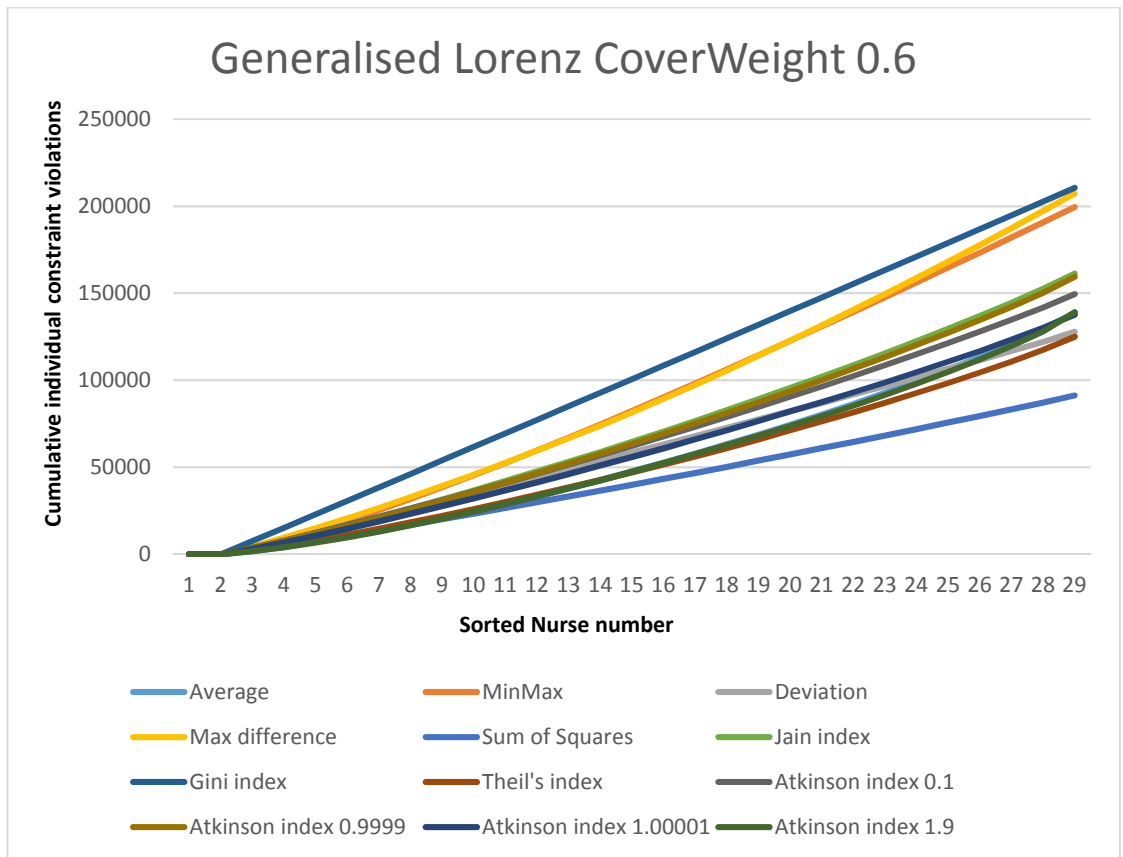
Figure 8-7: Generalised Lorenz CoverWeight 0.6

The results for nurse satisfaction weights set to 0.4 are more intersecting than the ones in the previous examples, that is normal since the importance of welfare drops. Gini index still is the one that is higher than any other measure at all points. Average crosses with 4 other measures (Atkinson index 1.9, Theils index, Square sum and minmax. Square sum crosses with Theil, Atkinson 1.9, minmax and Average. Max difference still is the second worse after Gini.
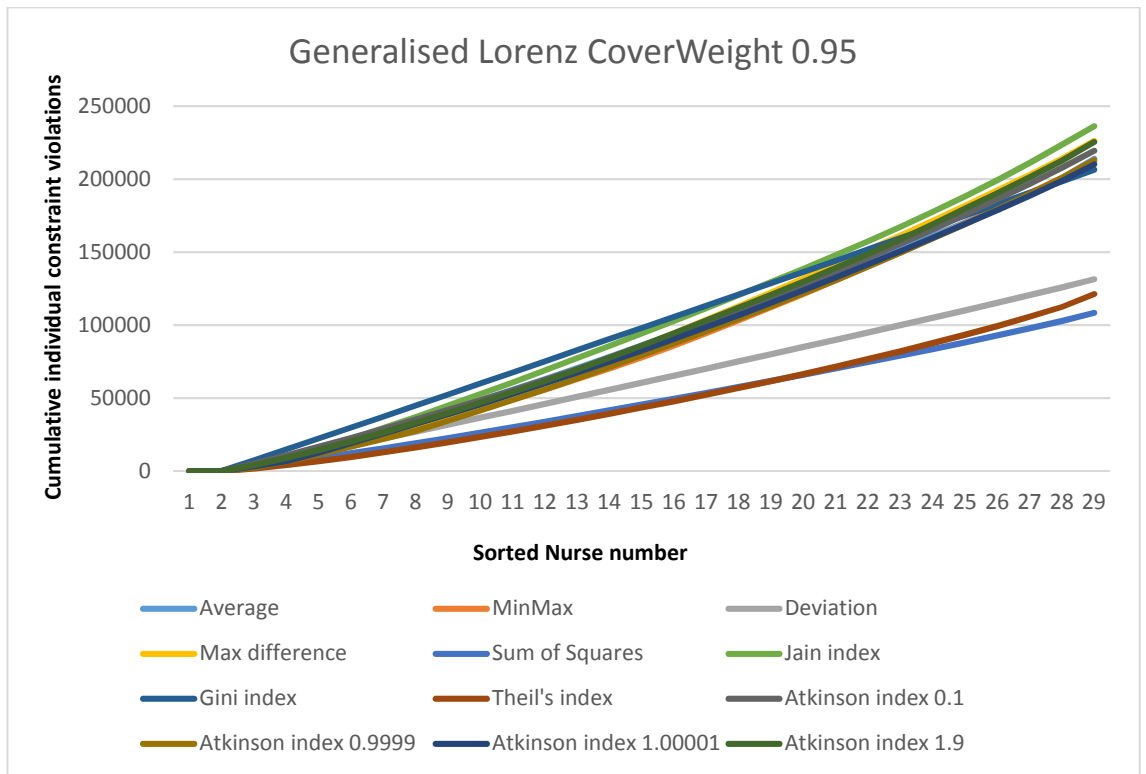
Figure 8-8

In the last example, where nurses' welfare is set to be almost neglectable, all GL curves intersect with another. Even Gini index that was clearly the worst in previous examples is crossing with Jain index. Average performance is between the 7th and 10th place, contrary to the other cases where average was among the best performing measures. It can be assumed that no safe conclusions can be drawn about nurses' welfare.

8.4.2.1 Case studies

In order to evaluate the sensitivity of some SWF to weight changes, comparison between the results of a SWF with different weights must be made. The GL curves of three selected measures are presented. It would be expected that for higher weights on nurse satisfaction the better (lower) the GL curves would be since minimizing constraint violations of nurses (and thus pointing the GL curve lower) would be more important.
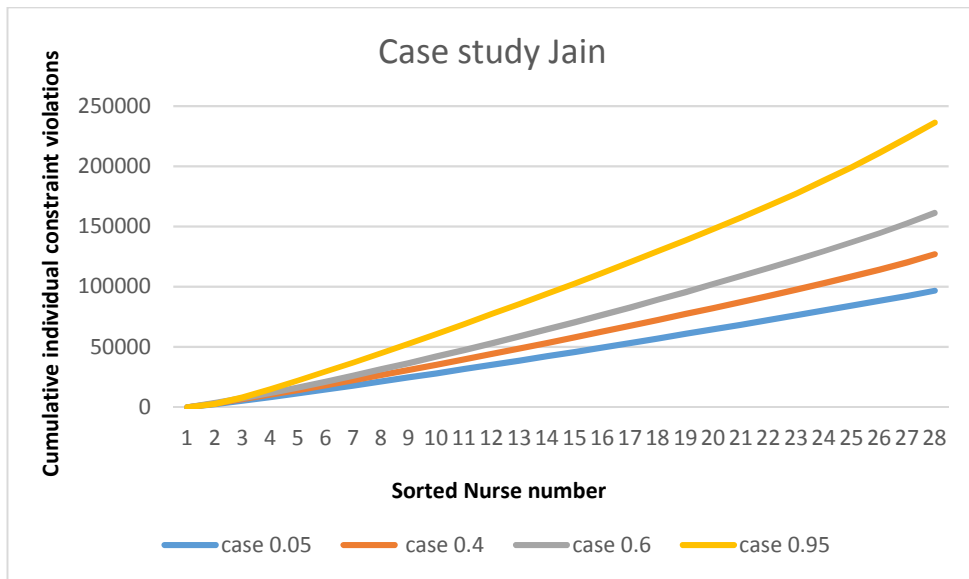
Figure 8-9: Case study Jain

Jain curves are the only completely consistent with our hypothesis. The higher the weight for nurses' satisfaction, the lower the curve. The curves do not intersect either.
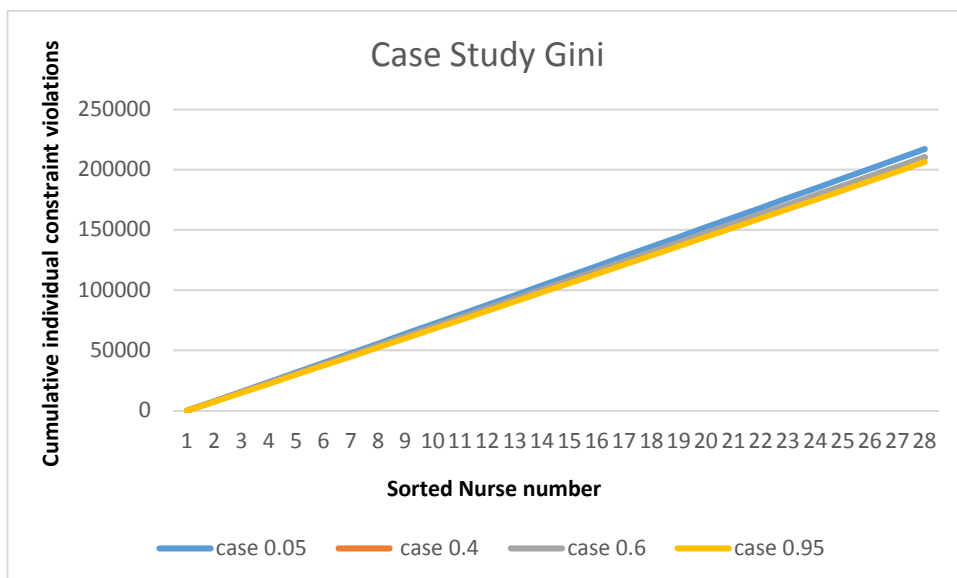
Figure 8-10: Case study Gini

Gini displays an opposite behavior: even though all curves are close and higher than the ones in the other studies (Jain and Theil) the order of results is 0.95 > 0.6> 0.4 >0.05.
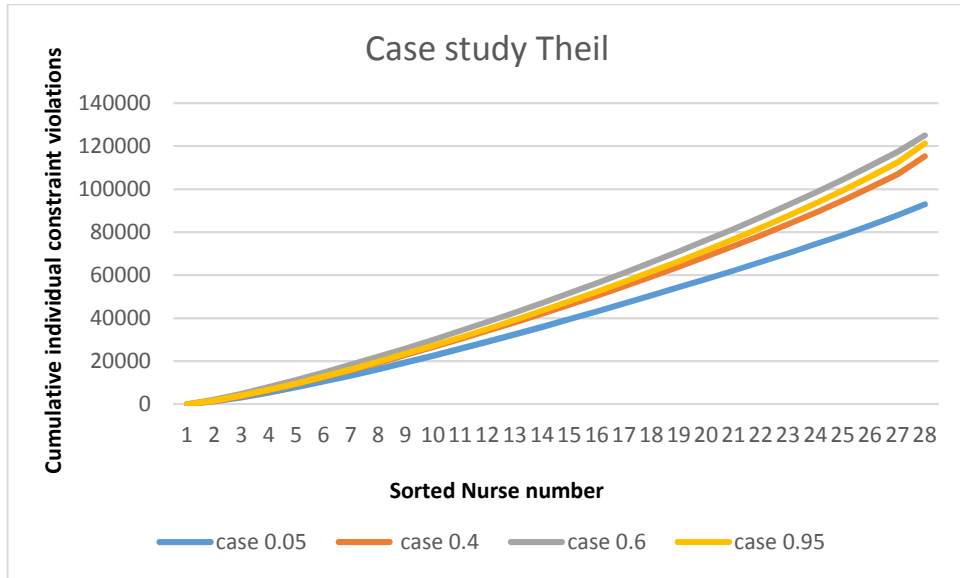
Figure 8-11: Case study Theil

Theil's index behaviour is somewhere in the middle. While case 0.05 is the best followed by 0.40, the two last ones cross.

From the above tables, the obvious remark that can be made is that Jain is the measure that is more sensitive to changes, Gini index seem to be the measure that is least sensitive, and Theil is somewhere in the middle. Jains results are the ones closer to expectations, following by Theil and Gini results are irrational.

8.4.3 ELECTRE

The results were evaluated by ELECTRE method (Roy 1991). The results from the different measures are the alternatives, while individual constraint violations are defined as criteria. Criteria are sorted, since the ones with more constraint violations are more important than the ones with less constraint violations. The weights for each criterion is given by this formula:

$$wt_i = \frac{(K - r_i + 1)^z}{\sum_{j=1}^{K}(K - j + 1)^z}$$

The value we chose for z equals to one. This ensures our goal for fairness, while still leaving space for efficiency.

Non-comparable (veto) conditions were set as well. Two solutions were deemed as non-comparable when both the following cases were true:

91

In any individual in the bottom 12.5% of the ranked table solution A was way worse than solution B. "Way worse" was defined as having more than triple constraint violation than its counterpart.

The total quality in solution A was found to be better than Solution B.

We state that there is a concordance threshold as a percentage of the weights that are in favour of one distribution against another. If while comparing distributions no single distribution exceeds the concordance threshold, we consider them to be indifferent towards each other. If one distribution exceeds the threshold, we can conclude that this distribution is better. We experimented with different concordance thresholds in those 3 cases.

Apart from nurse roster quality, other aspects of the results should be examined. Coverage results are of importance for a schedule to be accepted by the wards. In order to compare the measures, total running times and average running time for each iteration should be considered. To accommodate that, coverage constraints were incorporated into ELECTRE method. The results presented in the next table are from first instance with social welfare value weight set at 0.4, coverage weight at 0.6 (the graphs 8.2, 8.6 from chapter 8.4).

In order to incorporate the coverage satisfaction into the evaluation coverage constraint violations were given 50% of the total weight. The other 50% was assigned to nurses according to the formula above (also described in 4.4.1)

Measures	Average	MinMax	Deviation	Max diff	Square sum	Jain	Gini	Theil's	Atk 0.1	Atk 0.9999	Atk 1.0001	Atk 1.9
Average	X	0.00	0.10	0.50	Incomp	0.10	0.00	0.17	0.10	0.10	0.60	0.11
MinMax	**1.00**	X	0.47	**0.79**	0.50	0.50	0.00	0.50	**1.00**	**1.00**	**1.00**	0.50
Deviation	**0.90**	0.53	X	Incomp	**1.00**	**0.89**	0.50	**0.95**	**0.90**	**0.90**	**0.90**	**0.92**
Max diff	0.50	0.21	Incomp	X	0.50	0.47	0.11	0.50	0.47	0.47	0.47	0.50
Square sum	Incomp	0.50	0.00	0.50	X	0.50	0.50	0.07	0.50	0.50	0.50	0.57
Jain	**0.90**	0.50	0.11	0.53	0.50	X	0.00	0.48	**0.79**	0.92	**0.93**	0.45
Gini	**1.00**	**1.00**	0.50	**0.89**	0.50	**1.00**	X	0.50	1.00	**1.00**	**1.00**	**1.00**
Theil's	**0.83**	0.50	0.05	0.50	**0.93**	0.52	0.50	X	0.52	0.50	0.52	0.50
Atk 0.1	**0.90**	0.00	0.10	0.53	0.50	0.21	0.00	0.48	X	0.39	**0.87**	0.45
Atk 0.9999	**0.90**	0.00	0.10	0.53	0.50	0.08	0.00	0.50	0.61	X	0.60	0.48
Atk 1.00001	0.40	0.00	0.10	0.53	0.50	0.07	0.00	0.48	0.14	0.40	X	0.47
Atk 1.9	**0.89**	0.50	0.08	0.50	0.43	0.55	0.00	0.50	0.55	0.52	0.53	X

Table 8-4: ELECTRE method

The threshold for concordance test was set to be 0.75. Selecting the threshold this high ensures that one alternative has both coverage constraints (50% of weights) and at least half of the nurses' criteria weights (25% of total weights) to support the dominance.

Some observations about the above table. Average, Max diff, Square sum, and Atk1.0001 were not dominated by any other alternative. In the case of MaxDiff, this is largely because of good performance in Coverage constraint violations. Average outperformed most other measures, with the exception of Atk 1.0001.

8.4.4 Computational cost

In order to have a clearer view in the quality of the results produced, the amount of times an algorithm has been run and the speed of this algorithm is important. If two results are comparable in roster quality, but one is lacking in speed per iteration or total number of iterations, then this one can be considered.

The computational cost for each algorithm run included not only the calculation of the measure, but other factors as well. Apart from the normalisation, calculating the constraint violations, simulated annealing moves and coverage constraint violations were calculated each time an algorithm run was performed. Those calculations took the majority of time that's why as in chapter 6 the running time does not agree with the algorithmic analysis given in chapter 4.

The number of runs was similar in any measure, minimising the importance of the iteration factor, and leading us to safer conclusions.

Objective function	Time required for each run
Average	0.434825928
MinMax	0.861559748
Deviation	0.231503282
Max difference	12.2681077
Sum of Squares	0.41738748
Jain index	0.394580085
Gini index	0.454176934
Theil's index	0.465473573
Atkinson index 0.1	0.390843229
Atkinson index 0.9999	0.42947806
Atkinson index 1.00001	0.424199308
Atkinson index 1.9	0.392506182

Table 8-5: Computational cost

8.5 Conclusions

In this chapter it is introduced a novel approach to Social welfare in nurse rostering, introducing new measures used in the SWF concept. The measures are applied using existing search methods in popular Nurse rostering benchmarks. The results were encouraging; New measures performed satisfactorily, in many instances. Depending on the instance, Jain index, cases of Atkinson and Theil index shone between them.

Those outranking methods provide a framework for evaluating distributions with respect in fairness. Being flexible and easily adjustable can emulate personal views from any type of decision maker, more or less inequality averse, or efficient hungry. Thus, it can be stated that this framework can be applied in most problems where the aspect of fairness and efficiency could be present.

Next steps include applying this framework and measures to other OR problems. An interesting direction would be to implement other similar concepts to fairness in nurse rostering, such as poverty-starvation of resources.

Chapter 9 Goal programming

9.1 Introduction

Chapter 6 presented the results on fairness, that showed that when used as an exclusive goal the outcome is not applicable. Poverty results were interesting, and applicable by themselves, but some of the times high inequality results. A way to combine both poverty, fairness, efficiency and the necessary satisfaction of coverage constraints would be Goal programming (Jones 2011), described in 4.5.

9.2 Methodology

Goal programming can be used in order to combine different goals. In nurse rostering, the equivalent of a goal is the satisfaction of an individual nurse. The benchmarks used in this thesis have minimum of 19 nurses, creating a system with at least 19 different goals. This would not be practical, and meta goals were instead investigated. A meta goal is defined as a property of the system that is not a specific goal, but rather a side effect from the actual goal.

In this case, satisfying the "efficiency" goal would imply that the average nurse constraint violation is below a point, while satisfying a goal would imply that a specific nurse' sum of constraint violations is below a point.

The four meta goals were defined as follows:

- Efficiency: The sum of the sum of Nurses constraint violations
- Poverty was defined as the headcount ratio, the percentage of nurses that their sum of constraint violations was below a threshold
- Coverage was defined as the sum of hospitals needs' constraint violations
- For fairness models the measures in Chapter 6 were used.

In order to be able to compare goals in a similar order of magnitude, the above criteria were normalised. Headcount ration did not need to be normalised since it ranges between 0 (when no one is considered poor) and 1 (when everyone is above poverty line). As fairness goals, normalised fairness measures were used for simplicity. Those measures are Jains, Gini, Theils and variations of Atkinson measures. However, some steps for the normalisation of the other meta goals were needed.

A series of tests were performed in order to find a reasonable array of values for both coverage and efficiency constraints for each instance. The first set of experiments was performed with having only efficiency and only coverage as goals. Thus, we established the absolute best efficiency and worst coverage, and worst efficiency and best coverage respectively. The second set of experiments was run with their respective weights being 0.001 and 0.999, defining a lower limit of best and worst values.

9.3 Models

In order to evaluate the quality of nurse rosters produced by goal programming models' experiments were applied are again on benchmarks from Bilgin 2008. Those benchmarks provided 4 different wards with 2 different scenarios: emergency, geriatrics, psychiatry and reception wards, with nurses having identical or different preferences. The wards vary in parameters such as number of nurses needed, scheduling period, number of shifts and skills categorisation. The total number of constrains differ in each ward, with geriatrics ward being the most trivial and Emegerncy being the most complicated one. In the cases of scenarios with nurses having variable preferences, the constraints are personalised, but there still is a contract that is common for a set of nurses. To avoid presenting an unnecessary number of results, the results from the personalised and identical preferences were combined.

The weights between the 4 different meta-objectives on our GP model were set on the following way:

	Weight	Weight	Weight	Weight	Weight	Weight	Weight	Weight	Weight
Coverage	0.7	0.1	0.1	0.1	0.4	0.2	0.2	0.2	0.25
Efficiency	0.1	0.7	0.1	0.1	0.2	0.4	0.2	0.2	0.25
Fairness	0.1	0.1	0.7	0.1	0.2	0.2	0.4	0.2	0.25
Poverty	0.1	0.1	0.1	0.7	0.2	0.2	0.2	0.4	0.25

Table 9-1: Weight grid

Those weights were selected in order to better represent the views of different decision makers. The higher the weight in each goal is, the more important this goal is to the decision maker. In the first four cases, the decision maker showed strong preference over one criterion. The next four showed a more balanced approach to the four meta-goals, while still having one preference over the others. The last set of weights is completely balanced, signifying complete preference equality between the four goals. The weights are resembling the ones from Jones 2011 method.

As seen, in the majority of situations the collective goals for nurse happiness was weighted more than wards' needs.

Since poverty, efficiency and fairness goals all address to the issue of nurse satisfaction, another way to view this problem is to those 3 goals can be combined into one goals: nurse satisfaction. That would be another presentation of the same problem, merging the 3 meta-goals into one. The number of experiments performed for each case of total weight assigned to coverage and nurses' satisfaction is given in the next table 9.2.

	Weight	Weight	Weight	Weight	Weight
Coverage	0.7	0.1	0.4	0.2	0.25
Nurse satisfaction	0.3	0.9	0.6	0.8	0.75
Experiments	1 case	3 cases	1 case	3 cases	1 case

Table 9-2: Aggregated weight grid

The weighted goal program was modeled as follows:

$$OF_{weighted} = w_{cov} * p_1 + w_{eff} * p_2 + w_{fair} * p_3 + w_{pov} * p_4$$

$$S.T: \frac{cov - cov_{best}}{cov_{worst} - cov_{best}} - p_1 = 0$$

$$\frac{eff - eff_{best}}{eff_{worst} - eff\_best} - p_2 = 0$$

$$fair - p_3 = 0$$

$$pov - p_4 = 0$$

$w_{cov}, w_{eff}, w_{fair}, w_{pov}$ are the respective weights for the goals, and p_i are the positive deviational variables. Since the optimal values for normalised efficiency, coverage, fairness and poverty meta goals are 0, any positive deviation is penalised. The objective function of the goal programming model is to minimise those deviations.

The Chebyshev formulation has the same constraints:

$$OF_{minmax} = \max(p_1, p_2, p_3\ p_4)$$

$$S.T: \frac{cov - cov_{best}}{cov_{worst} - cov_{best}} - p_1 = 0$$

$$\frac{eff - eff_{best}}{eff_{worst} - eff\_best} - p_2 = 0$$

$$fair - p_3 = 0$$

$$pov - p_4 = 0$$

The objective function of the Chebyshev formulation is to minimise the maximum deviation from the meta-goals.

"Fair" is defined by the model used, be it either Jain, Gini, Theil or Atkinson.

9.4 Results

Since the number of possible combinations between models, instances and weights is a large one, the results between problems with identical and different preferences were combined. So, the presented results are on four benchmarks instead of 8. This is made in order to have a clearer presentation of the results.

The results are going to be presented in two ways. Firstly, results for one particular instance and for particular measure, in order to explore the differences between results from different weights. Secondly, results for one instance, and a particular view of the central decision maker weights will be presented, with the results from all measures. This way the measures will be compared by weight value.

9.5 Evaluation

Evaluating the results of an allocation is a problem by itself. the models are based on individual happiness as well as goal satisfaction, so a cardinal evaluation is really hard. Using cardinal measures such as Jain in Martin 2013 is not an option, since those measures are either already used, or could have been used in the formulation. Thus, an evaluation based on individual happiness is required to evaluate the results. Generalised Lorenz Curves, satisfaction levels and a form of "fair ELECTRE" were applied.

Generalised Lorenz curves are preferred in this case since they represent both the distribution and the utility that is distributed.

In order to present the results in a clearer way an analysis of the measure is performed. A selection of the best measure is made in order to enable a comparison against each measure.

Satisfaction levels

Another way to view evaluate results is by satisfaction level tables. In those tables the normalised performance of each criterion is shown. Poverty and fairness are always within the limits of zero and one, however in cases where coverage and efficiency weights are low,

these criteria might be out of those limits. This can be explained by the levels of ideal and anti-ideal values selected.

The antithesis between covering hospitals' needs and satisfying nurses is clear. In the cases where efficiency was the most important factor coverage criterion was more than twice the allowed max. That can be explained of the direct opposition between those two: having low nurses' satisfaction constraint violations is antithetical to having low hospitals constraint violations. The antithesis between fairness and efficiency is apparent when fairness is weighted really high: having a fair roster do not equal to having an efficient one. This agrees with the results from chapter 8. Efficiency and starvation of resources seem to have a synergy: when weight in poverty criterion is high, efficiency criterion has a low value as well. This makes sense, since the goal of poverty criterion is to push individuals lower than a point, while the goal of efficiency is to push individuals as low as possible.

Another interesting viewpoint is the differences in trade-offs different fairness measures bring. Atkinson and Gini index are quite "strong", and that leads to a complete contrast with efficiency: in most cases where fairness is weighted 0.7 and Gini and Atkinson measures are examined, efficiency is more than the least desired threshold. The lower achievements in efficiency usually lead to worse results in poverty as well. In every case where Fairness was more important than other measures weighting either 70% or 40% of total weights, poverty and efficiency were underperforming.

9.5.1 Lorenz curves

Lorenz curves were formulated using only the individual happiness results, and excluding the coverage results. The reason behind that is the way lays in the Lorenz curves formulation: it is included individual happiness and excluded the overage score, since it would need more assumptions on how coverage score is compared with individuals. For example, in cases where coverage constrains had a bigger score than the worst individual coverage should be the first observation. However, if coverage is smaller than the best individual it should be in the last part of the curve. A compromise could be that the coverage should always be the first observation of the curve making it always more important than any individual, or last making it always the least important. This would go beyond the point of weighting the trade-offs between different goals.

The curves presented here are from the combined problems 1 (Emergency) and 4 (Geriatrics). Emergency instance have the largest number of nurses (27), and geriatrics have 19 (smallest along with reception, but with less number of shift types. It is worth noting that both Reception and Psychiatry (aggregated instances 2 and 3) display very similar behaviour with Emergency in all measures.

Problem1:
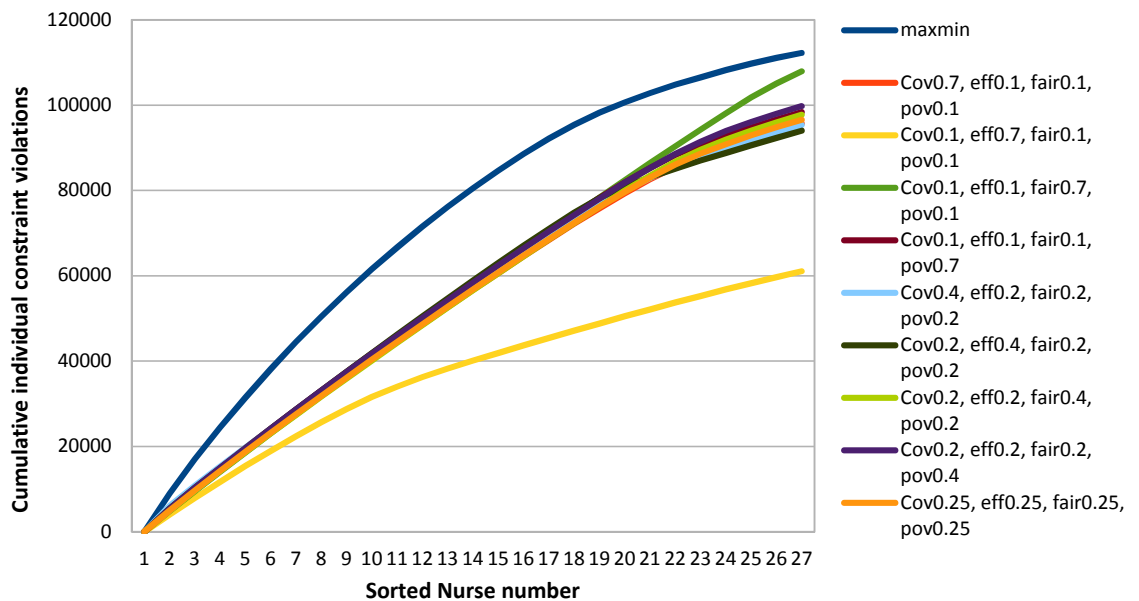
Figures 9-1 Atkinson Lorenz curve, aggregated problem 1

The GL curve for GP with Atkinson measure shows that the best (lower) curve is the one produced by assigning high weights in efficiency (0.7 efficiency weight). The worst results are the ones produced by minmax. The rest of the curves are crossing, with the curve for fairness 0.7 being second lowest in the first nurses and second highest in the last ones. This is to be explained since fairness generally gives greater priority to the first nurses, the ones that have most constraint violations. The results produced by fairness 0.4 start being 4th lowest, to 2nd lowest to 6th lowest. That is a more balanced approach of the fairness 0.7, that do not perform as good at the nurses with lower constraint violations or raise as high in the nurses with the most constraint violations.

The first results presented are from Atkinson measure.
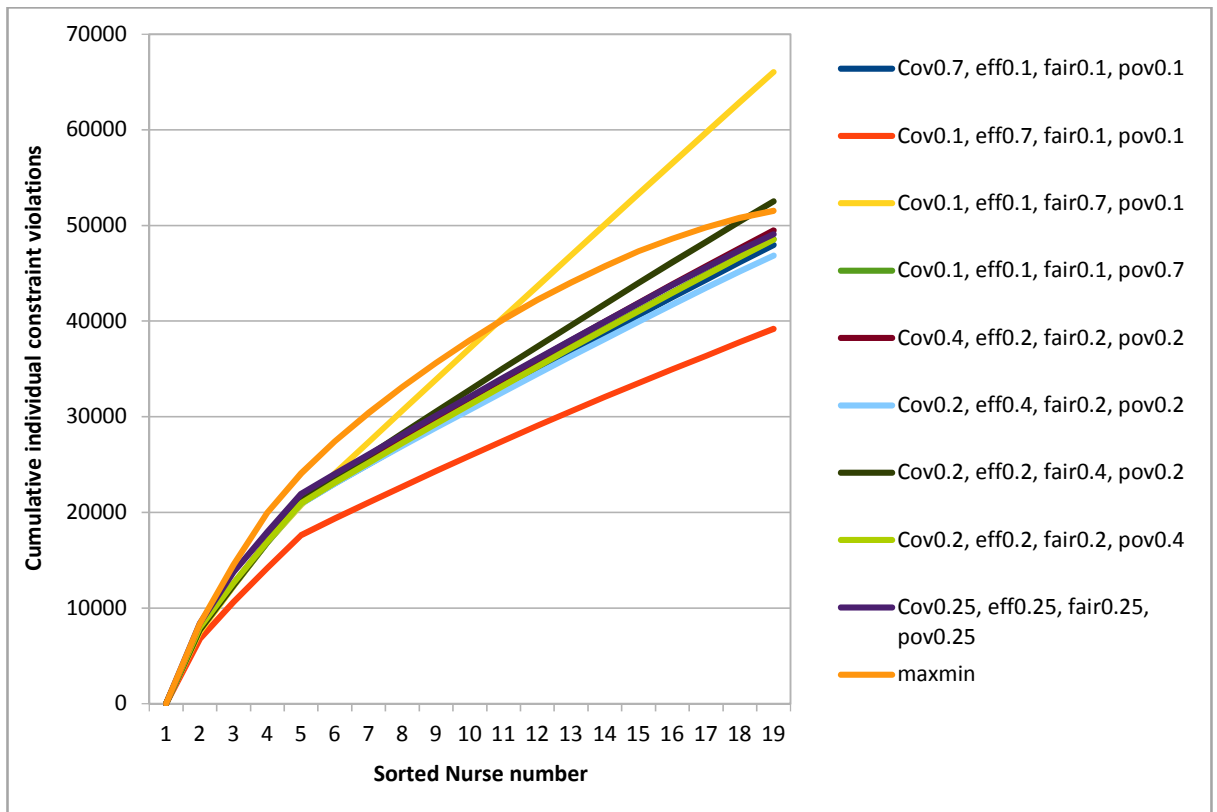
Figure 9-2: Atkinson Lorenz curve, aggregated problem 4

The results in instance 4 (Geriatrics) are quite different than the other instances. The results produced by high coverage (coverage 0.7) are the lowest at all points. Minmax provide the 3rd-4th best results. High efficiency starts as the second-best curve for the nurse with most constraint violations and ends up being the highest. It is also worth noting that most curves cross.
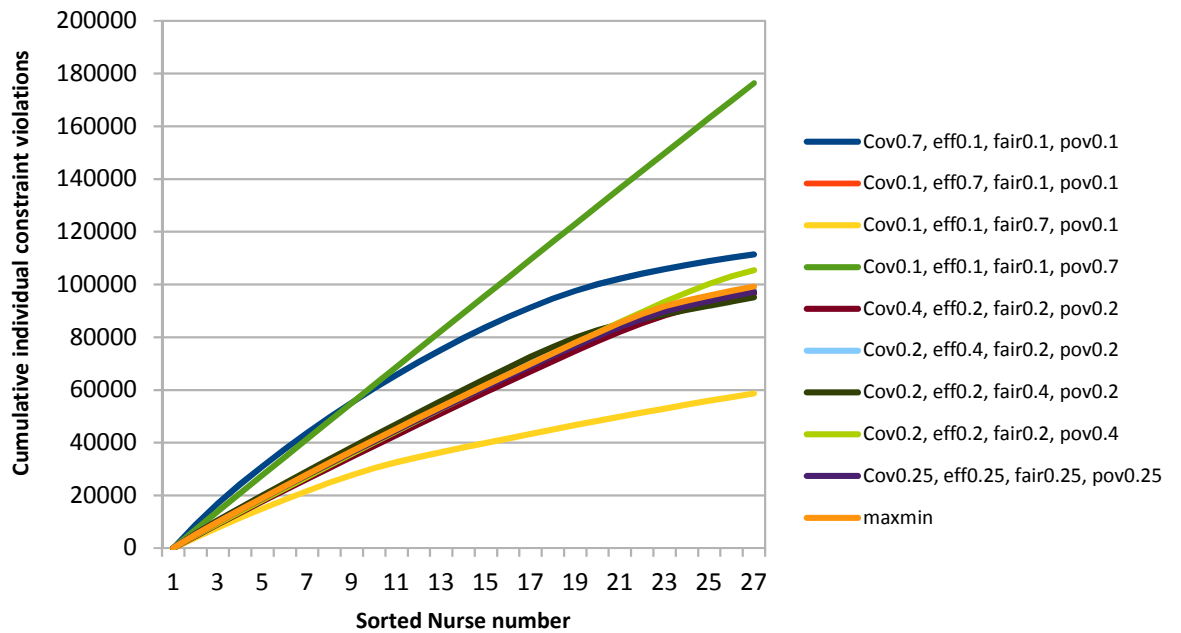
Figure 9-1: Gini Lorenz curve, aggregated problem 1

Results from Gini are somewhat following the pattern from the results produced by Atkinson. The Lorenz curve produced by high efficiency (efficiency 0.7) is the lowest without crossing with any other curve. The curve produced by high poverty (poverty 0.7) are second lowest till cumulative nurse number 21, only then outperformed by second high efficiency (efficiency 0.4) that in turn underperformed (with a high curve) in the first nurses. Results produced by high fairness and minmax were the ones with the highest curves, outperformed by any other measure.
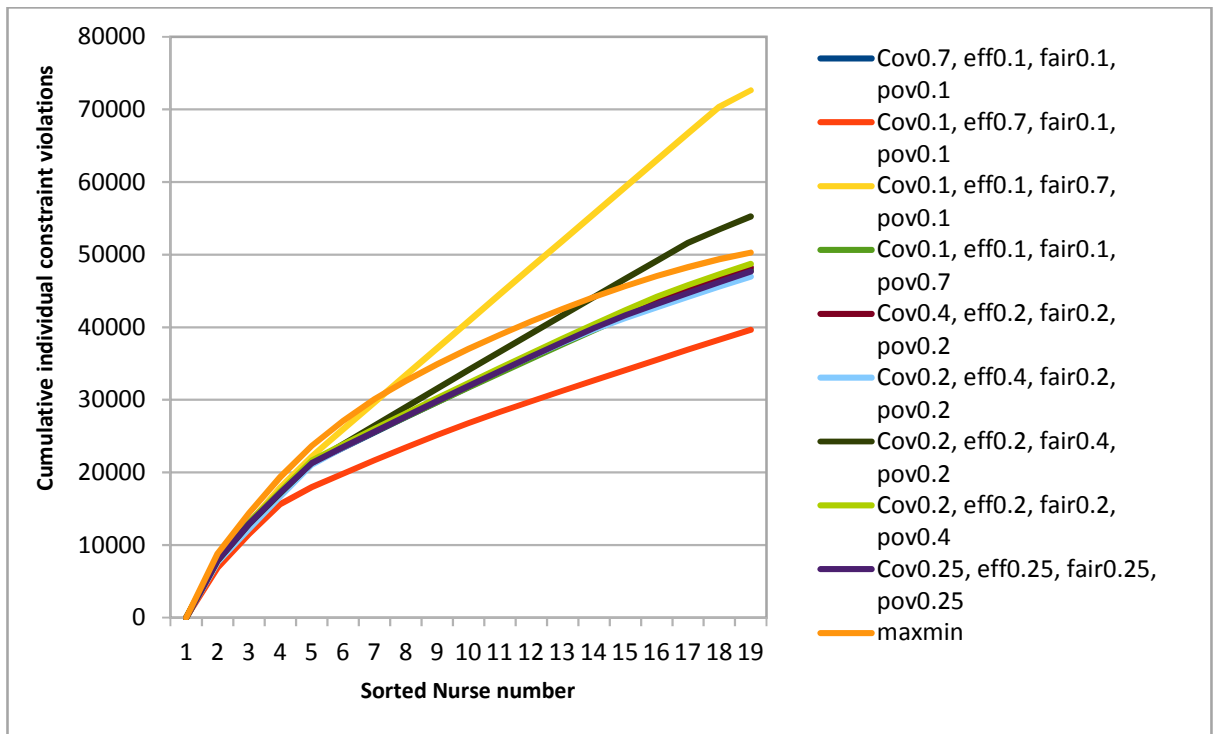
Figure 9-2 Gini Lorenz curve, aggregated problem 4

As in Atkinson, the best result produced was by the high coverage (coverage 0.7). The results produced by the second highest coverage (coverage 0.4) performed were second lowest in the start and the end of the curve, while dropping up to the 6th place at around the middle of the curve. Minmax also ranged between 3rd and 5th. Apart from the first nurse, the curve of high efficiency (efficiency 0.7) was outperformed by all other results, being higher than the rest.
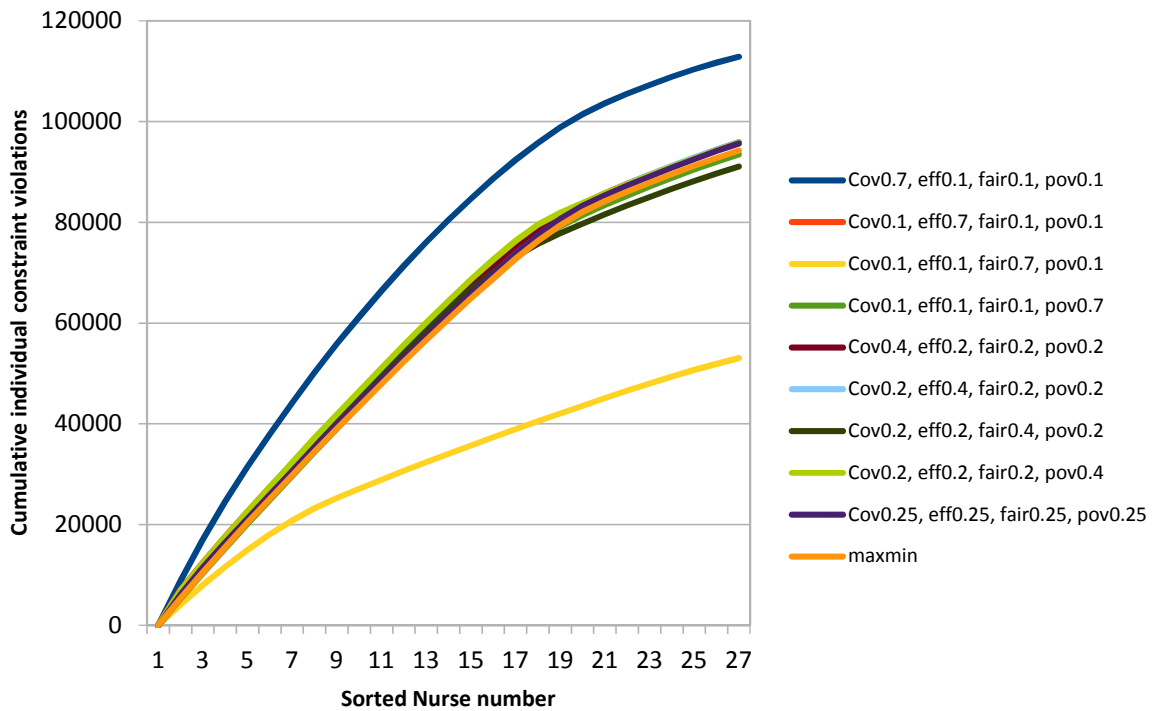
Figure 9-3: Jain Lorenz curve, aggregated problem 1

As in the previous graphs, high efficiency (efficiency 0.7) curve was the lowest at all points. High fairness (fairness 0.7) was the second lowest at the first nurses, and ranged between 2nd and 4th best. Balanced results (0.25) ranged between 3rd and 4th best. Somewhat balanced fairness focused (fairness 0.4) started well as the 4th lowest curve, dropping to 7th lowest and then being the second lowest curve in the nurses 18-27. Minmax again was thehighest curve, without crossing with any other.
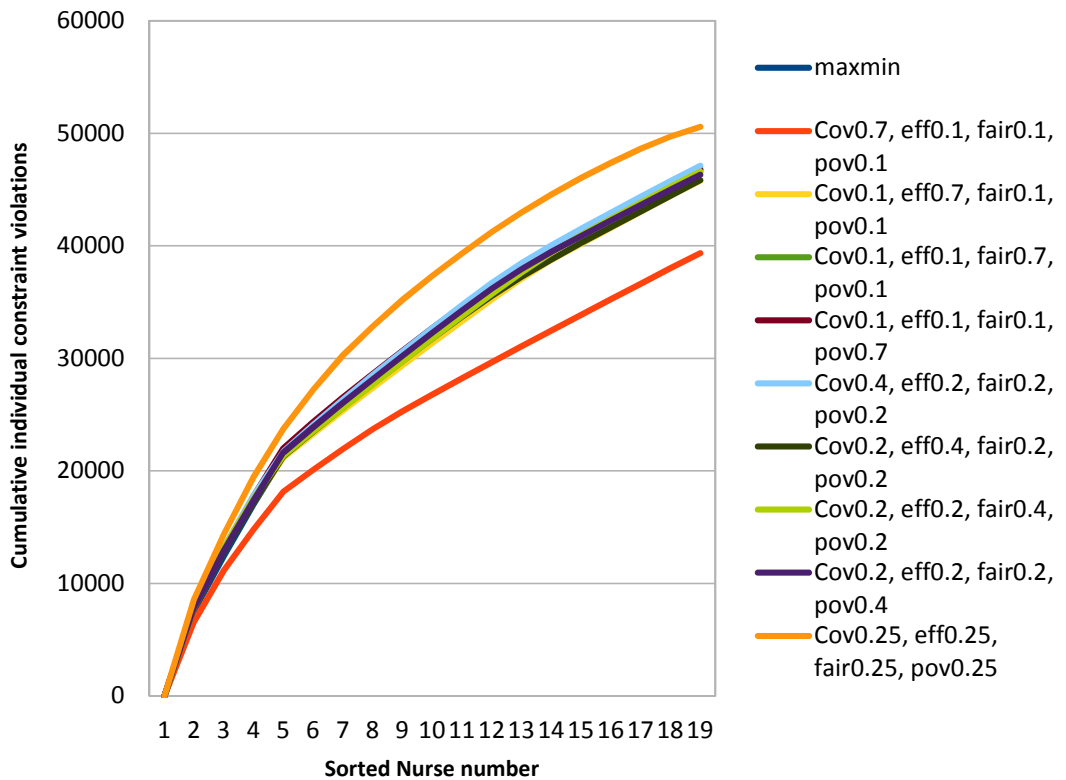
Figure 9-4: Jain Lorenz curve, aggregated problem 4

High coverage (coverage 0.7) curve was the lowest, without crossing with any other curve. Balanced weights (0.25) was the highest at all points without crossing with any other curve. High efficiency (0.7) started as a normal performing curve (4th -5th) but from cumulative nurse number 4 was the second lowest curve, remaining that way till the end. Minmax performed well early (2nd) but declined at the later nurses (8th). High fairness did not achieve something better than the 5th lower curve. Somewhat balanced efficiency ranged from 3rd to 4th lower curve.
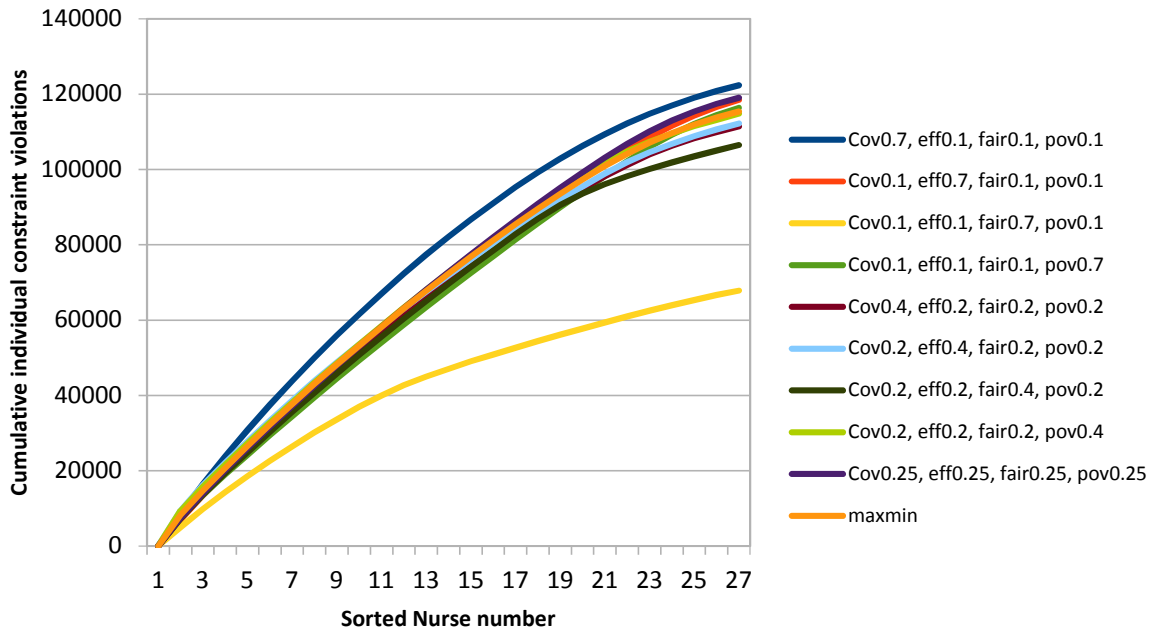
Figure 9-5: Theil's Lorenz curve, aggregated problem 1

Unsurprisingly, High efficiency curve was the lowest at all points. Relatively high weight efficiency ranged from 2nd best to 4th best. High poverty performed bad at the first nurses, and improved being 3rd best. Minmax was at the last at almost all points.
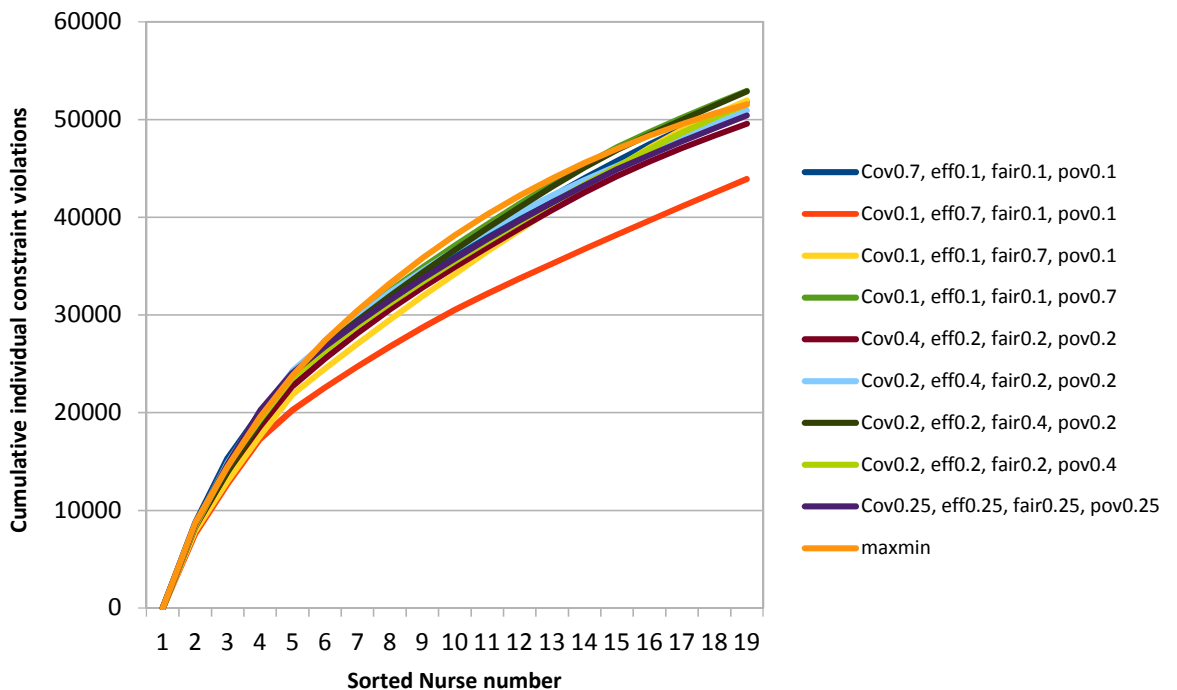
Figure 9-6: Theil Lorenz curve, aggregated problem 4

Finally, in the combined geriatrics instance, high coverage was the lowest curve. High efficiency performed well in the first nurses but dropped to be the 8th best. Moderately high

poverty performed bad at the worst-off nurses, but at the latest points of the curve was the 3rd lower. Moderately high efficiency and high fairness started relatively well but ended up being among the highest curves.

The results in Lorenz curves are mixed. In the majority of the problems, the best result is when efficiency criterion is weighted highly. In the aggregated 4th instance the dominating results are the ones from the coverage criterion. It would be expected to have models that weight efficiency highly performing well. It would make sense in a way that those cases are trying to improve all individuals, something that is easily illustrated in Lorenz curves. Measures with high weights in fairness would be expected to perform better in the first part of the curve, where the most worse-off nurses are shown, and less well in the latter part of the curve. This is the case in three out of four of those combined instances.

Some possible explanations could be that the ideal and anti-ideal points set were not representative of the instances. Selecting ideal and anti-ideal points affects the performance of the objective functions. For example, if an anti-ideal point is set too low, some objective function might appear to perform well in that meta goal. If an ideal point is set too high, an objective function might appear to perform bad in this meta goal.

Another reason for the difference of the performance on different weights of the meta goals could be that the difference between constraint violations for nurses in instances. Having a different number of constraint violations might affect the difference between ideal and anti-ideal points and consequently the meta-goals performance.

The difference between the number of nurses might also influence meta goals. For example, having a smaller number of individuals might make a fairness measure much more sensitive in changes.

9.5.2 Satisfaction Levels

The satisfaction levels of the four novel measures are presented. Since all meta-goals are either bounded (fairness, poverty) or normalised (efficiency, coverage) it would be natural to view the performance of each objective function. The purpose of satisfaction levels is to show the performance of each objective function for each meta goal, ranged between 0 and 1. Values close to zero would represent the achievement of near ideal point values for this particular meta goal. In the table with weighted formulation, there are some objective functions with values that surpass 1 in some meta goals. While in the result generation this value would be equal to 1, in the following table numbers larger than 1 represent

performance worse than the anti-ideal point at this particular meta-goal. This happened because even after experiments to identify ideal and anti-ideal points, some objective functions valued some meta-goals low, in a way that allowed for worse performance than anti-ideal points.

Table 9.3 presents the satisfaction of the minmax formulation.

Problem0	Coverage	Efficiency	Fairness	Poverty
Jain	0.69426	0.050979	0.685185	0.694293
Gini	0.663847	0.322581	0.666667	0.649758
Theil	0.795126	0.034593	0.796296	0.790912
Atkinson	0.724305	0.519183	0.703704	0.721467
Problem2	Coverage	Efficiency	Fairness	Poverty
Jain	0.863809	0.067896	0.857143	0.802885
Gini	0.85453	0.302985	0.857143	0.826923
Theil	0.850048	0.071551	0.833333	0.658654
Atkinson	0.903524	0.45672	0.904762	0.903846
Problem4	Coverage	Efficiency	Fairness	Poverty
Jain	0.877759	0.070264	0.736842	0.879808
Gini	0.830985	0.298981	0.736842	0.807692
Theil	0.861403	0.058491	0.763158	0.783654
Atkinson	0.802353	0.593829	0.763158	0.759615
Problem6	Coverage	Efficiency	Fairness	Poverty
Jain	0.915753	0.080238	0.868421	0.893275
Gini	0.945562	0.422339	0.868421	0.947368
Theil	0.972959	0.091077	0.894737	0.973684
Atkinson	0.940917	0.753081	0.921053	0.945906

Table 9-3: Satisfaction levels for all instances: min max formulation

In the min max formulation, Atkinson and in a lesser extend Gini have the most balanced results, with relative smaller differences between the performance for the different meta goals. Efficiency is generally not that important when trying to minimise the maximum meta goal since the criteria that have high values are coverage, fairness and poverty.

Table 9.4 presents the results from the weighted formulation for instance 3 (combined reception instances).

Cov	Eff	Fair	Pov			Cov	Eff	Fair	Pov	
0.1	1	0.2	0.8	7/1/1/1	Atk	0	0.9	0.1	0.7	7/1/1/1	Jain
3.4	0.1	0.1	0.2	1/7/1/1		3.8	0	0.1	0.2	1/7/1/1
0.1	1.2	0	1	1/1/7/1		0.2	0.9	0.1	0.8	1/1/7/1
0.1	0.9	0.1	0.9	1/1/1/7		0.1	0.9	0.1	0.7	1/1/1/7
0.1	0.9	0.1	0.9	4/2/2/2		0.1	0.9	0.1	0.8	4/2/2/2
0.1	0.9	0.1	0.8	2/4/2/2		0.2	0.9	0.1	0.7	2/4/2/2
0.1	1	0.1	0.8	2/2/4/2		0.2	0.9	0.1	0.7	2/2/4/2
0.1	1	0.1	0.9	2/2/2/4		0.1	0.9	0.1	0.8	2/2/2/4
0.1	0.9	0.1	0.9	2.5		0.1	0.9	0.1	0.7	2.5
0.8	0.6	0.8	0.8	minmax		0.9	0.1	0.7	0.9	minmax

Cov	Eff	Fair	Pov			x	Eff	Fair	Pov	
0.1	1	0.1	0.8	7/1/1/1	Gini	0	0.9	0.1	0.7	7/1/1/1	Theil
3.6	0	0.2	0.2	1/7/1/1		3.8	0	0.1	0.2	1/7/1/1
0.1	1.6	0	1	1/1/7/1		0.2	0.9	0.1	0.8	1/1/7/1
0.1	1	0.2	0.8	1/1/1/7		0.1	0.9	0.1	0.7	1/1/1/7
0.1	1	0.1	0.8	4/2/2/2		0.1	0.9	0.1	0.8	4/2/2/2
0.1	0.9	0.2	0.8	2/4/2/2		0.2	0.9	0.1	0.7	2/4/2/2
0.2	1.1	0.1	0.8	2/2/4/2		0.2	0.9	0.1	0.7	2/2/4/2
0.1	1	0.2	0.8	2/2/2/4		0.1	0.9	0.1	0.8	2/2/2/4
0.1	1	0.1	0.9	2.5		0.1	0.9	0.1	0.7	2.5
0.8	0.3	0.7	0.8	minmax		0.9	0.1	0.7	0.9	minmax

Table 9-4: Satisfaction levels for problem 4

Generally, performance of the different objective functions for fairness is good with the exception of minmax, where fairness is relatively higher than the rest of the weights. Values over 1 appear in coverage performance, and in a lesser extend in efficiency. Poverty and fairness are already bounded meta-goals. Poverty meta-goal responds good to higher weights for efficiency. The explanation between poverty efficiency possible correlation is that while poverty "pushes" for a lower point, the goal of efficiency is to push as low as possible. Efficiency and coverage are completely antithetical, with the worst performance for coverage coming when efficiency weight is set high and vice versa.

9.5.3 ELECTRE

As it can be noted, the best results between models are incomparable. In order to compare the results between different measures "fair" ELECTRE was applied. The weights were selected in an effort to compromise between ward needs and nurse's needs. The weight of

the coverage constrain violation was set to be 50%, and the weight for each individual was given by this function:

$$w_i = \frac{(N - r_i + 1)^e}{\sum_1^n (N - r_i + 1)^e}$$

The results of the ELECTRE show a slight preference to the measures that resulted in lower coverage constraint violation. This can be subject to change, depending on the weights given to the different criteria. The threshold for concordance was set to be 0.7 in order to make sure that dominance is not given to an objective function without this objective function being better than another at the coverage meta goal, and at least 40% of the weights of nurses.

	Atk	Gini	Jain	Theil
Atk	X	0.50	0.50	**0.77**
Gini	0.50	X	0.49	0.69
Jain	0.50	0.51	X	0.50
Theil	0.23	0.31	0.50	X

Table 9-5: ELECTRE aggregated problem 0

A preference for Theil's measure against Atkinson's measure is the only definite result in the first instance.

	Atk	Gini	Jain	Theil
Atk	X	**0.91**	0.48	0.39
Gini	0.09	X	0.50	0.20
Jain	0.52	0.50	X	0.00
Theil	0.61	**0.80**	**1.00**	X

Table 9-6: ELECTRE aggregated problem 2

A preference of Gini and Jain over Theil, and Gini over Atkinson was shown in the second instance.

	Atk	Gini	Jain	Theil
Atk	X	0.50	0.50	**0.84**
Gini	0.50	X	0.44	0.59
Jain	0.50	0.56	X	0.50
Theil	0.16	0.41	0.50	X

Table 9-7: ELECTRE aggregated problem 4

A preference of Theil over Atkinson was shown in the third instance

	Atk	Gini	Jain	Theil
Atk	X	0.3038277512	0.4043062201	0.2631578947
Gini	0.6961722488	X	0.4282296651	**0.7511961722**
Jain	0.5956937799	0.5717703349	X	0.6052631579
Theil	**0.7368421053**	**0.7488038278**	0.3947368421	X

Table 9-8: ELECTRE aggregated problem 6

Theil and Gini had the same Coverage constraints, so they Were equally good. Atkinson was preferred to Theil in the fourth instance.

Generally, the objective functions that perform well in nurses' satisfaction (see table 9.2) underperform in coverage satisfaction. A lot of pairwise comparisons are close to the 0.5 point because of the above. Clear evidence of the dominance of one solution over the other would require dominance on coverage constraint, and usually on the worse of nurses. In this aspect, and because of the negative correlation between efficiency and coverage (see 9.5.2) it is sometimes possible for objective functions that underperform in efficiency to show dominance over objective functions that do not. Selecting different weight for coverage and nurses satisfaction, or even changing the formula that the weight for nurse satisfaction is given would see differences in the ELECTRE comparisons.

9.6 Conclusions

In this dissertation a novel approach to evaluating societal problems has been introduced. A Goal program including new meta-goals was introduced. Different ways of evaluating results on a cardinal basis, using ELECTRE and graphical methods were applied in a large set of results.

Trade-offs and dependencies between four different goals in a specific problem were hinted. Relative strength of different fairness measures towards different goals was tested.

Chapter 10 conclusions

The main goal of this thesis is to create the framework and prove the viability of fairness related solutions in operational research (OR) problems . The need for those goals stems from the lack of an established framework for inequality averse solutions in the context of decision making, and the limited application of fairness objective functions in OR problems (Chapter 3).

Chapter 5 discusses the properties and concepts of poverty, or minimisation of unhappiness. This is an area of economics that is neglected in the science of decision making. To the writer's knowledge, it is the first time that poverty – as a concept taken from economics- is applied in an operational research problem. This chapter tries to introduced poverty in operational research and establish a framework about poverty. Insight for use of poverty in goal programming is provided, and the possible connection between GP and poverty. The goal of this chapter is to show that poverty measurement can be used as an objective in OR problems.

Fairness models are tested in chapter 6. Experiments are made using models presented in previous chapters. The results show that even though the fairness model produce equal rosters the rosters are so ineffective that would not be preferred by most of the individuals involved. Comparing rosters for each individual nurse shows that having fair rosters does not mean they are also effective. This proves the claims of Kaplow and Shavel (2003) and suggests that in order for fairness to be used, it must be part of a series of goals, or part of the objective function instead.

Poverty and minimisation of dissatisfaction is examined in chapter 7. Poverty measurement and identification is discussed. Results show that like fairness, poverty averse rosters are not necessarily effective. Non-poverty measures such as Jain and Average seem to do more to alleviate poverty than some poverty measures. Also from the experiments it was shown that the importance of choosing the appropriate poverty line is of some importance: defining the poverty affects the perceived happiness or unhappiness. However, aside the perceived happiness for nurses, poverty alone did not produce good results in terms of effectiveness or improving the status of the less "privileged" nurses. However, some simpler measures like poverty gap did achieve decent results in the aspect of targeting the poverty line, leaving room for improvement in future experiments.

In chapter 8 the Social Welfare Function (SWF) was introduced. SWF combined the concepts of fairness and efficiency in one objective function. In experiments the results are improved in fairness (compared to efficiency only measures) and in regards to efficiency (compared to fairness only measures). A weight sensitivity analysis was also performed, showing tradeoffs between nurse happiness and hospital needs.

Goal programming model was used in Chapter 9. Objectives from previous chapters, namely chapter 6 and 7 were combined with efficiency measure and coverage constraints. This created a powerful tool that allows flexibility in decision making. The results show that, depending on the preferences, good results can be obtained on all desired criteria.

To summarise, the thesis contributions were

- An introduction of fairness models, and a Framework for fairness in OR.
- An introduction of poverty, poverty models and a framework for poverty in OR.
- Proof of concept that poverty and fairness can be used, with certain limitations, in decision making.
- Proof of concept that poverty and fairness are viable goals in a set of Goal programming and lead to interesting results.
- Social welfare functions are leading to exceptional results.

10.1 Future directions

Fairness and poverty in OR is an interesting topic with vast potential applications. The work in this thesis focused in a particular problem of operational research (nurse rostering), in order to analyse fairness, poverty, SWF, and a goal programming formulation comprising of fairness and poverty. However, this can be expanded to other OR problems that involve individuals and distributable resources. Other methodologies could be included as well. Some possible applications would be the following:

- Apply poverty and fairness in different areas of OR, such as facility location problem.
- Adjust search function, or even poverty measures to be more compatible
- Include more/different poverty/fairness goals in the goal programming model formulation.
- Investigate the feasibility of those methods in problems with a larger number of individual entities

- Expand Goal Programming using poverty concepts, for example see Jones 2017.

Finally, an interesting direction for future research would be to view the implications of applying fairness, SWF, Poverty or a combination of those objectives in a real-world application, such as an established economy. The effects of interactions between inequality or poverty aversion and effectiveness, as defined by Gross domestic products would be of importance. An interesting application of could be Greece: a country where less resources are distributed every passing day, and with a raising inequality and poverty indexes. The question in hand is: can we use fairness, poverty or a combination of those measures in order to decide what allocation of resources is the best – and thus enforce reforms that would benefit the weaker people of a country in crisis?

References

Alfares, H. K., & Duffuaa, S. O. (2008). Assigning cardinal weights in multi-criteria decision making based on ordinal ranking. Journal of Multi-Criteria Decision Analysis, 15(5-6), 125-133.

Alkire, Sabina, and Séverine Deneulin. "Introducing the human development and capability approach." An introduction to the human development and capability approach. London: Earthscan (2009).

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. Journal of public economics, 95(7), 476-487.

Alkire, Sabina, and Maria Emma Santos. "Acute multidimensional poverty: A new index for developing countries." (2010).

Amartya Sen, "Equality of What?" in A. Sen, Choice, Welfare and Measurement, (Oxford: Blackwell, 1982).

Anand, Sudhir. "Aspects of poverty in Malaysia." Review of Income and Wealth 23, no. 1 (1977): 1-16.

Arrow, Kenneth J. "A difficulty in the concept of social welfare." Journal of political economy 58, no. 4 (1950): 328-346.

Arulambalam, Ambalavanar, Xiaoqiang Chen, and Nirwin Ansari. "Allocating fair rates for available bit rate service in ATM networks." IEEE Communications magazine 34, no. 11 (1996): 92-100.

Atkinson, Anthony B. "On the measurement of inequality." Journal of economic theory 2, no. 3 (1970): 244-263.

Atkinson, A. B. (1970). Poverty in Britain and the reform of social security (No. 18). Cambridge University Press.

Atkinson, A. B. (1987). On the measurement of poverty. Econometrica: Journal of the Econometric Society, 749-764.

Atkinson, A. B. (2003). Multidimensional deprivation: contrasting social welfare and counting approaches. The Journal of Economic Inequality, 1(1), 51-65.

Atkinson, Anthony B. "On the measurement of inequality." Journal of economic theory 2.3 (1970): 244-263.

Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. Management Science, 42(11), 1515-1523.

Batchelder,A. B. (I97i). The Economics of Poverty. New York: Wiley.

Beckerman, W. (1979). 'The impact of income maintenance payments on poverty, 1975.' ECONOMIC JOURNAL, vol. 86, pp. 261-79.

Beach, C.M. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. The Review of Economic Studies 50, 723-35.

Belleschi, Marco, Gabor Fodor, and Andrea Abrardo. "Performance analysis of a distributed resource allocation scheme for D2D communications." In GLOBECOM Workshops (GC Wkshps), 2011 IEEE, pp. 358-362. IEEE, 2011.

Bentham, J. (1879). An introduction to the principles of morals and legislation. Clarendon Press.

Bergson, A. (1938). A reformulation of certain aspects of welfare economics. The Quarterly Journal of Economics, 310-334.

Berman, Oded, and Edward H. Kaplan. "Equity maximizing facility location schemes." Transportation Science 24, no. 2 (1990): 137-144.

Berrada, Ilham, Jacques A. Ferland, and Philippe Michelon. "A multi-objective approach to nurse scheduling with both hard and soft constraints." Socio-Economic Planning Sciences 30.3 (1996): 183-193.

Bertsimas, Dimitris, and Shubham Gupta. "A proposal for network air traffic flow management incorporating fairness and airline collaboration." Operations Research (2011).

Bertsimas, Dimitris, and Sarah Stock Patterson. "The air traffic flow management problem with enroute capacities." Operations research 46, no. 3 (1998): 406-422.

Bertsimas, Dimitris, Vivek F. Farias, and Nikolaos Trichakis. "Fairness, efficiency, and flexibility in organ allocation for kidney transplantation." Operations Research 61, no. 1 (2013): 73-87.

Bertsimas, D., Farias, V. F., & Trichakis, N. (2011). The price of fairness. Operations research, 59(1), 17-31.

Bilgin, B. "Project web page of automation of nurse rostering in Belgian hospitals." (2008) https://people.cs.kuleuven.be/~pieter.smet/nurserostering.html

Bilgin, Burak, Patrick De Causmaecker, Benoît Rossie, and Greet Vanden Berghe. "Local search neighbourhoods for dealing with a novel nurse rostering model." Annals of Operations Research 194, no. 1 (2012): 33-57.

Bilgin, Burak, Patrick De Causmaecker, Benoit Rossie, and Greet Vanden Berghe. "The problem description and a solution method for the nurse rostering problem in Belgian hospitals." In Proceedings of the 22nd Annual Conference of the Belgian Operations Research Society (ORBEL 22), vol. 22, pp. 37-38. 2008.

Blackburn, M. L. (1989). Poverty measurement: an index related to a Theil measure of inequality. Journal of Business & Economic Statistics, 7(4), 475-481.

Blegen, M. A. (1993). Nurses' job satisfaction: a meta-analysis of related variables. Nursing research, 42(1), 36-41.

Bonald, Thomas, and Laurent Massoulié. "Impact of fairness on Internet performance." In ACM SIGMETRICS Performance Evaluation Review, vol. 29, no. 1, pp. 82-91. ACM, 2001.

Booth, Charles. Life and Labour of the People in London. Vol. 8. Macmillan and Company, 1903.

Booth, Charles, ed. Labour and Life of the People... Vol. 1. Williams and Norgate, 1891.

Bourguignon, Francois, and Satya R. Chakravarty. "The measurement of multidimensional poverty." Journal of Economic inequality 1, no. 1 (2003): 25-49.

Bourguignon, Francois. "Decomposable income inequality measures." Econometrica: Journal of the Econometric Society (1979): 901-920.

Bradshaw, Jonathan, and Emese Mayhew. "The measurement of extreme poverty in the European Union." (2011).

Bronfenbrenner, M. [1973], "Equality and equity", Annals of the American Academy of Political and Social Science, 409, 1, pp. 9–23.

Burke, Edmund K., Patrick De Causmaecker, Greet Vanden Berghe, and Hendrik Van Landeghem. "The state of the art of nurse rostering." Journal of scheduling 7, no. 6 (2004): 441-499.

Burke E. K. , P. De Causmaecker, S. Petrovic, and G. Vanden Berghe. Fitness evaluation for nurse scheduling problems. In Proceedings of the Congress on Evolutionary Computation (CEC2001), pages 1139–1146, Seoul, Korea, May 27-30 2001. IEEE Press.

Burke, Edmund K., Tim Curtois, Rong Qu, and Greet Vanden Berghe. "Problem model for nurse rostering benchmark instances." ASAP, School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham, UK (2008).

Burke, Edmund, Peter Cowling, Patrick De Causmaecker, and Greet Vanden Berghe. "A memetic approach to the nurse rostering problem." Applied intelligence 15, no. 3 (2001): 199-214.

Burke, Edmund, Patrick De Causmaecker, and Greet Vanden Berghe. "A hybrid tabu search algorithm for the nurse rostering problem." Simulated evolution and learning. Springer Berlin Heidelberg, 1999. 187-194.

Cheang, Brenda, Haibing Li, Andrew Lim, and Brian Rodrigues. "Nurse rostering problems—a bibliographic survey." European Journal of Operational Research 151, no. 3 (2003): 447-460.

Church, Richard, and Charles R. Velle. "The maximal covering location problem." Papers in regional science 32, no. 1 (1974): 101-118.

Cohen, G. A., "On the Currency of Egalitarian Justice," Ethics 99 (1989).

Constantino, Ademir A., Everton Luiz de Melo, Dario Landa-Silva, and Wesley Romão. "A heuristic algorithm for nurse scheduling with balanced preference satisfaction." In Computational Intelligence in Scheduling (SCIS), 2011 IEEE Symposium on, pp. 39-45. IEEE, 2011.

Dagum, Camilo. "On the relationship between income inequality measures and social welfare functions." Journal of Econometrics 43.1 (1990): 91-102.

Dalton, H. (1920). The measurement of the inequality of incomes. The Economic Journal, 30(119), 348-361.

Dalton, H. [1920], "The Measurement of the inequality of incomes", The Economic

Daly, M., Wilson, M., & Vasdev, S. (2001). Income inequality and homicide rates in Canada and the United States. Canadian J. Criminology, 43, 219.

Dasgupta, P. [1993], An inquiry into well-being and destitution, Oxford University

Datt, G., & Ravallion, M. (1992). Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s. Journal of development economics, 38(2), 275-295.

De Bruin, A. M., Bekker, R., Van Zanten, L., & Koole, G. M. (2010). Dimensioning hospital wards using the Erlang loss model. Annals of Operational research, 178(1), 23-43.

Deaton, A. (1997). The analysis of household surveys: a microeconometric approach to development policy. World Bank Publications.

Demery, Lionel, and David Demery. "Poverty and macroeconomic policy in Malaysia, 1979–1987." *World development* 19, no. 11 (1991): 1615-1632.

Demery, L., & Squire, L. (1996). Macroeconomic adjustment and poverty in Africa: an emerging picture. The World Bank Research Observer, 11(1), 39-59.

Deutsch, M. [1975], "Equity, equality, and need: what determines which value will be used as the basis of distributive justice?", Journal of Social Issues, 31, 3, pp. 137–149.

Dijkstra, E. W. "Solution of a problem in concurrent programming control."Communications of the ACM 8, no. 9 (1965): 569.

Donaldson, David, and John A. Weymark. "Properties of fixed-population poverty indices." International Economic Review (1986): 667-688.

Drezner, T., Drezner, Z., & Guyse, J. (2009). Equitable service by a facility: Minimizing the Gini coefficient. Computers & Operational research, 36(12), 3240-3246.

Drougas, Y., & Kalogeraki, V. (2005, March). A fair resource allocation algorithm for peer-to-peer overlays. In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE (Vol. 4, pp. 2853-2858). IEEE.

Dworkin, R. [1981], "What is equality? Part 1: wquality of welfare", Philosophy and Public Affairs, 10, 3, pp. 185–246.

Egbelu, Pius J., and Jose MA Tanchoco. "Characterization of automatic guided vehicle dispatching rules." The International Journal of Production Research 22, no. 3 (1984): 359-374.

Ernst T. A, H. Jiang, M. Krishnamoorthy, B. Owens and D. Sier (2004). "Annotated Bibliography of Personnel Scheduling and Rostering." Annals of Operational research, Volume 127, pages 21-144.

Eryilmaz, A., & Srikant, R. (2005, March). Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. In Proceedings IEEE 24th Annual Joint Conference of the IEEE Com

Ferrera, M. (Ed.). (2005). Welfare state reform in southern Europe: Fighting poverty and social exclusion in Greece, Italy, Spain and Portugal. Routledge.

Fishburn, P. C. (1977): "Mean-Risk Analysis with Risk Associated with Below-Target Returns," American Economic Review, 67, 116-126.

Foster, J. E., & Shorrocks, A. F. (1988a). Poverty orderings and welfare dominance. In Distributive Justice and Inequality (pp. 91-110). Springer Berlin Heidelberg.

Foster, J. E., & Shorrocks, A. F. (1988b). Poverty orderings. Econometrica: Journal of the Econometric Society, 173-177.

Foster, J. E., & Shorrocks, A. F. (1991). Subgroup consistent poverty indices. Econometrica: Journal of the Econometric Society, 687-709.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. Econometrica: Journal of the Econometric Society, 761-766.

Fuchs, V. (1967). Redefining poverty and redistributing income. The Public Interest, (8), 88.

Gastwirth, J. L. (1972). The estimation of the Lorenz curve and Gini index. The review of economics and statistics, 306-316.

Gastwirth, J.L. and M. Gail (1985). Simple asymptotically distribution-free methods for comparing Lorenz curves and Gini indices obtained from complete data. In R. Basmann and G. Rhodes (Eds.), Advances in Econometrics, Volume 4, 229-43. Greenwich, CT: JAI Press.

Gini, C. (1912). "Italian: Variabilità e mutabilità" (Variability and Mutability', C. Cuppini, Bologna, 156 pages. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955).

Glasser, G.J. (1962). Variance formulas for the mean difference and coefficient of concentration. Journal of the American Statistical Association 57, 648-54.

Glassey, C. R., & Resende, M. G. (1988). A scheduling rule for job release in semiconductor fabrication. Operational research Letters, 7(5), 213-217.

Goedhart, T., Halberstadt, V., Kapteyn, A., & Van Praag, B. (1977). The poverty line: Concept and measurement. Journal of Human Resources, 503-520.

Gold, E. Mark. "Deadlock prediction: Easy and difficult cases." SIAM Journal on Computing 7, no. 3 (1978): 320-336.

Gopalan, R., & Batta, R. (1990). The equity constrained shortest path problem.Computers & Operational research, 17(3), 297-307.

M Ul Haq, Mahbub, and I. Kaul. "Human development report." United Nations Development Programme (1990).

Hart, J. F. (1978). Computer approximations. Krieger Publishing Co., Inc..

Harter, James K., Frank L. Schmidt, and Theodore L. Hayes. "Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis." Journal of applied psychology 87.2 (2002): 268.

Haughton, J., & Khandker, S. R. (2009). Handbook on poverty+ inequality. World Bank Publications.

Hayes, B., Bonner, A. N. N., & Pryor, J. (2010). Factors contributing to nurse job satisfaction in the acute hospital setting: a review of recent literature.Journal of Nursing Management, 18(7), 804-814.+

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics 19, 293-325.

Hoemmen, Mark. "Computing the standard deviation efficiently." (2007).

Jain, R., Chiu, D. M., & Hawe, W. R. (1984). A quantitative measure of fairness and discrimination for resource allocation in shared computer system. Eastern Research Laboratory, Digital Equipment Corporation.

Jalan, J., & Ravallion, M. (1998). Transient poverty in postreform rural China.Journal of Comparative Economics, 26(2), 338-357.

Jeremy Bentham (1989). The Collected Works of Jeremy Bentham: Correspondence: Volume 9: January 1817 to June 1820.

Joe-Wong, C., Sen, S., Lan, T., & Chiang, M. (2012, March). Multi-resource allocation: Fairness-efficiency tradeoffs in a unifying framework. In INFOCOM, 2012 Proceedings IEEE (pp. 1206-1214). IEEE.

Jones, D. (2011). A practical weight sensitivity algorithm for goal and multiple objective programming. European Journal of Operational Research,213(1), 238-245.

Jones, D., & Tamiz, M. (2010). Practical goal programming (Vol. 141). New York: Springer.

Kakwani, N. (I980). 'On a class of poverty measures.' Econometrica, vol. 48, PP. 437-46.

Kaplow , Louis, and Steven Shavell. Fairness versus welfare: notes on the Pareto principle, preferences, and distributive justice. No. w9622. National Bureau of Economic Research, 2003.

Kaplow, Louis, and Steven Shavell. "Fairness versus welfare." Harv. l. rev.114 (2000): 961.

Kaplow, Louis, and Steven Shavell. "The conflict between notions of fairness and the Pareto principle." American Law and Economics Review 1, no. 1 (1999): 63-77.

Katajainen, J., & Träff, J. L. (1997, March). A meticulous analysis of mergesort programs. In Italian Conference on Algorithms and Complexity (pp. 217-228). Springer, Berlin, Heidelberg.

Kilpatrick, R. W. (1973). The income elasticity of the poverty line. The Review of Economics and statistics, 327-332.

Kim, Kap Hwan, Su Min Jeon, and Kwang Ryel Ryu. "Deadlock prevention for automated guided vehicles in automated container terminals." InContainer Terminals and Cargo Systems, pp. 243-263. Springer Berlin Heidelberg, 2007.

Knuth, D. E. (1998). The art of computer programming: sorting and searching (Vol. 3). Pearson Education.

Kovner, C., Brewer, C., Wu, Y. W., Cheng, Y., & Suzuki, M. (2006). Factors associated with work satisfaction of registered nurses. Journal of Nursing Scholarship, 38(1), 71-79.

Lambert, P. (2001). The distribution and redistribution of income. Manchester University Press.

Lan, T., Kao, D., Chiang, M., & Sabharwal, A. (2010). An axiomatic theory of fairness in network resource allocation (pp. 1-9). IEEE.

Leach, J. (2009). Income Disparity, Inequity Aversion and the Design of the Healthcare System. The Scandinavian Journal of Economics, 111(2), 277-297.

Le-Anh*, Tuan, and M. B. M. De Koster. "On-line dispatching rules for vehicle-based internal transport systems." International Journal of Production Research 43, no. 8 (2005): 1711-1728.

Le-Anh, Tuan, and J. Robert van der Meer. "Testing and classifying vehicle dispatching rules in three real-world settings." Journal of Operations Management 22, no. 4 (2004): 369-386.

Lehmann, Matthias, Martin Grunow, and Hans-Otto Günther. "Deadlock handling for real-time control of AGVs at automated container terminals." Or Spectrum 28, no. 4 (2006): 631-657.

Lerman, R. I., & Yitzhaki, S. (1985). Income inequality effects by income source: a new approach and applications to the United States. The review of economics and statistics, 151-156.

Li J-A, Leung SCH, Wu Y, Liu K (2007) Allocation of empty containers between multi-ports. Eur J Oper Res 182:400–412

Li J-A, Liu K, Leung SCH, Lai KK (2004) Empty container management in a port with long-run average criterion. Math Comput Model 40:85–100

Lim, Jae Kook, Kap Hwan Kim, Kazuho Yoshimoto, Jun Ho Lee, and Teruo Takahashi. "A dispatching method for automated guided vehicles by using a bidding concept." In Container Terminals and Automated Transport Systems, pp. 325-344. Springer Berlin Heidelberg, 2005.

Litzkow, M. J., Livny, M., & Mutka, M. W. (1988). Condor- a hunter of idle workstations. In Distributed Computing Systems, 1988., 8th International Conference on (pp. 104-111). IEEE.

Lorenz, Max O. "Methods of measuring the concentration of wealth." Publications of the American Statistical Association 9.70 (1905): 209-219.

De Grano, Melanie L., D. J. Medeiros, and David Eitel. "Accommodating individual preferences in nurse scheduling via auctions and optimization." Health Care Management Science 12, no. 3 (2009): 228.

M. Warner. Nurse staffing, scheduling, and reallocation in the hospital. Hospital & Health Services Administration, pages 77–90, 1976.

Marsh, M. T., & Schilling, D. A. (1994). Equity measurement in facility location analysis: a review and framework. European Journal of Operational Research,74(1), 1-17.

Martin, Simon, Djamila Ouelhadj, Pieter Smet, Greet Vanden Berghe, and Ender ÖZcan. "Cooperative search for fair nurse rosters." Expert Systems with Applications 40, no. 16 (2013): 6674-6683.

Marx, Karl. "Capital, volume I." (1867).

Marx, Karl. "Critique of the Gotha Program", Selected Works of Karl Marx and Frederick Engels, Eng. ed., FLPH, Moscow, 1951, Vol. 2, p. 23

Matsaganis, M., & Leventi, C. (2013). The distributional impact of the Greek crisis in 2010. Fiscal Studies, 34(1), 83-108.

Matsaganis, Manos. "The welfare state and the crisis: the case of Greece." Journal of European Social Policy 21, no. 5 (2011): 501-512.

McAllister, D. M. (1976). Equity and efficiency in public facility location. Geographical Analysis, 8(1), 47-63..

Milanovic, B. (1997). A simple way to calculate the Gini coefficient, and some implications. Economics Letters, 56(1), 45-49.

Moorthy, Rajeeva Lochana, Wee Hock-Guan, Ng Wing-Cheong, and Teo Chung-Piaw. "Cyclic deadlock prediction and avoidance for zone-controlled AGV system." International Journal of Production Economics 83, no. 3 (2003): 309-324.

Mueller, C. W., & McCloskey, J. C. (1990). Nurses' job satisfaction: a proposed measure. Nursing research, 39(2), 113-116.

Mühlenthaler, M., & Wanka, R. (2013). Fairness in academic course timetabling. arXiv preprint arXiv:1303.2860.)

Nash Jr, J. F. (1950b). The bargaining problem. Econometrica: Journal of the Econometric Society, 155-162.

Nash, J. F. (1950a). Equilibrium points in n-person games. Proceedings of the national academy of sciences, 36(1), 48-49.

Orshansky, M. (1965). Counting the poor: Another look at the poverty profile.Soc. Sec. Bull., 28, 3.

Ouelhadj, D., Martin, S., Smet, P., Ozcan, E., & Vanden Berghe, G. (2012). Fairness in nurse rostering.

Pyatt, G., Chen, C. N., & Fei, J. (1980). The distribution of income by factor components. The quarterly journal of economics, 95(3), 451-473.

Radunovic, B., & Le Boudec, J. Y. (2007). A unified framework for max-min and min-max fairness with applications. Networking, IEEE/ACM Transactions on,15(5), 1073-1083.

Rafaeli, A., Kedmi, E., Vashdi, D., & Barron, G. (2003). Queues and fairness: A multiple study experimental investigation.

Rainwater, L. (1974). What money buys: Inequality and the social meanings of income. Basic Books (AZ).

Ravallion, M. (1996). Issues in measuring and modelling poverty. The Economic Journal, 106(438), 1328-1343.

Rawls, J. [1971], A Theory of justice, Harvard University Press, Cambridge, USA.

Richard J. Arneson, "Equality and Equal Opportunity for Welfare," Philosophical Studies 56 (1989)

Ringuest, J. L., Graves, S. B., & Case, R. H. (2004). Mean–Gini analysis in R&D portfolio selection. European Journal of Operational Research, 154(1), 157-169.

Robb, Elizabeth A., Amy C. Determan, Lucy R. Lampat, and Mary J. Scherbring. "Self-scheduling: Satisfaction guaranteed?." Nursing Management 34, no. 7 (2003): 16.

Romero, C., 2001. Extended lexicographic goal programming: a unifying approach. Omega 29, 63–71

Romero, C., 2004. A general structure of achievement function for a goal programming model. European Journal of Operational Research 153, 675–686. Saaty, T.L., 1980. The Analytical Hierarchy Process. McGraw-Hill, New York.

Ronald Dworkin, Sovereign Virtue: The Theory and Practice of Equality (Cambridge, MA: Harvard University Press, 2000

Rönnberg, Elina, and Torbjörn Larsson. "Automating the self-scheduling process of nurses in Swedish healthcare: a pilot study." Health Care Management Science 13.1 (2010): 35-53.

Rowntree, B. S. (1901). Poverty: A study of town life. Macmillan.

Roy, A. D. (1952): "Safety First and the Holding of Assets," Econometrica, 20, 431-449.

Roy, B. (1968). Classement et choix en présence de points de vue multiples.RAIRO-Operational research-Recherche Opérationnelle, 2(V1), 57-75.

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. Theory and decision, 31(1), 49-73.

Rutstein, Shea Oscar, Kiersten Johnson, and ORC Macro MEASURE. *The DHS wealth index*. ORC Macro, MEASURE DHS, 2004.

Schechtman, E. and S. Yitzhaki (1987). A measure of association based on Gini's mean difference. Communications in Statistics – Theory and Methods 16, 207-31

Sen A. 1976: Poverty: an Ordinal Approach to Measurement, Econometrica, 44, 219-231

Sen, A 1970 (a) Collective Choice and Social Welfare

Sen, A. (1970) (b). The impossibility of a Paretian liberal. The journal of political economy, 78(1), 152-157.

Sen, A. (1979). Issues in the Measurement of Poverty. The Scandinavian Journal of Economics, 285-307.

Sen, A. (1980). Equality of what?. The Tanner lectures on human values, 1, 353-369.

Sen, A. (Ed.). (1973). On economic inequality. Oxford University Press.

Sendler, W. (1979). On statistical inference in concentration measurement. Metrika 26, 109-22.

Shin, H., & Lee, J. H. (2010, May). Subcarrier allocation for multiuser two-way OFDMA relay networks with fairness constraints. In Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st (pp. 1-5). IEEE.

Shorrocks, Anthony F. "Ranking income distributions." Economica (1983): 3-17.Tang, J., Hincapié, R., Xue, G., Zhang, W., & Bustamante, R. (2010). Fair bandwidth allocation in wireless mesh networks with cognitive radios. Vehicular Technology, IEEE Transactions on, 59(3), 1487-1496.

Smet, P., Bilgin, B., De Causmaecker, P., & Berghe, G. V. (2014). Modelling and evaluation issues in nurse rostering. Annals of Operations Research, 218(1), 303-326.

Smith, Adam. The wealth of nations [1776]. na, 1937.

Stillwell, W. G., Seaver, D. A., & Edwards, W. (1981). A comparison of weight approximation techniques in multiattribute utility decision making. Organizational behavior and human performance, 28(1), 62-77.

Tang, J., Hincapié, R., Xue, G., Zhang, W., & Bustamante, R. (2010). Fair bandwidth allocation in wireless mesh networks with cognitive radios. Vehicular Technology, IEEE Transactions on, 59(3), 1487-1496.

Theil, H., & Theil, H. (1967). Economics and information theory (No. 04; HB74. M3, T4.).

Townsend, P. (1954). Measuring poverty. The British Journal of Sociology, 5(2), 130-137.

Townsend, P. (1974). Poverty as relative deprivation: resources and style of Living. Wedderburn (ed.), Poverty, Inequality and Class Structure.

Trichakis, N. K. (2011). Fairness in operations: from theory to practice (Doctoral dissertation, Massachusetts Institute of Technology).

Tsakloglou, P. (1990). Aspects of poverty in Greece. Review of Income and Wealth, 36(4), 381-402.

Vanhoucke, Mario, and Broos Maenhout. "NSPLib—a nurse scheduling problem library: a tool to evaluate (meta-) heuristic procedures." Operational research for health policy: making better decisions. Proceedings of the 31st annual meeting of the working group on operational research applied to health services. 2007.

Vasupongayya, S., & Chiang, S. H. (2005, November). On Job Fairness in Non-Preemptive Parallel Job Scheduling. In IASTED PDCS (pp. 100-105).

Wang, Sheng-Pen, Yu-Kuang Hsieh, Zheng-Yun Zhuang, and Nai-Chia Ou. "Solving an outpatient nurse scheduling problem by binary goal programming." Journal of Industrial and Production Engineering 31, no. 1 (2014): 41-50.

Warner, D. M. (1976). Scheduling nursing personnel according to nursing preference: A mathematical

Warner, D. M., & Prawda, J. (1972). A mathematical programming model for scheduling nursing personnel in a hospital. Management Science, 19(4-Part-1), 411-422.

WATTS, H. W. (1968): "An Economic Definition of Poverty," in On Understanding Poverty, ed. by D. P. Moynihan. New York: Basic Books.

Wolfe, Harvey. A multiple assignment model for staffing nursing units. Diss. Johns Hopkins University, 1964.

World Bank Group (Ed.). (2012). World Development Indicators 2012. World Bank Publications.

Yang, M. (2011). Resource Allocation to Improve Equity in Service Operations(Doctoral dissertation, University of Cincinnati).

Yapicioglu, H., & Smith, A. E. (2011, June). Disservice representation using the Gini coefficient in semi-desirable facility location problems. In Evolutionary Computation (CEC), 2011 IEEE Congress on (pp. 110-114). IEEE.

Zangaro, G. A., & Johantgen, M. (2009). Registered nurses' job satisfaction in Navy hospitals. Military Medicine, 174(1), 076-081.

Zheng, Buhong. "Aggregate poverty measures." Journal of economic surveys 11, no. 2 (1997): 123-162.