FURTHERING THE INTEGRATION OF EQUITY AND OPTIMIZATION: THE BIN PACKING PROBLEM

ZANE SHOWALTER-CASTORENA

"it has to start somewhere, it has to start sometime, what better place than here, what better time then now?"
-Zack de la Rocha

1. Introduction

Before presenting the central focus of this paper it is necessary to familiarize readers with two preliminary concepts; equity and operations research. Equity is defined by Presidential Executive Order as "The consistent and systematic fair, just, and impartial treatment of all individuals, including individuals who belong to underserved communities that have been denied such treatment." [Exec. Order No. 13648, 2021]. More simply, it is the notion that certain populations require more resources than others in order to pursue life, liberty, and happiness. Diametrically opposed to this idea is the rigorous and dispassionate mathematics of Operations Research, which is a discipline of math that uses optimization tools to make a 'best' decision given specific parameters, variables, and a defined objective. It has been said the whole is greater than the sum of its parts, we believe this principal applies here as the goal of this paper is to demonstrate how equity can be incorporated into the well known bin packing problem. Specifically, we replace the objective function in the traditional integer programming formulation with the Kolm Polak Equally Distributed Equivalent (KP-EDE).

A equally distributed equivalent is a way to measure inequality in a distribution. Originally developed in an economic context to measure inequality in wealth distribution [Atkinson, 1970] EDE's measure both the level of wealth for an individual (or group) and the variation of wealth in the total population. As an example, consider the following situation:

- All Members/groups of a population are required to distribute various amounts of money to separate, necessary, services in order to survive.
- This spend puts them at a specific level of risk.
- This risk level is not uniform in the population, due to savings, increased earnings, etc.

The EDE is the level of risk at which any/all members of that population would become indifferent to a choice between a risk distribution in which everyone receives that same risk, and the current unequal risk distribution.

This idea is more fully motivated in future sections. We next motivate Operations Research, as it is the mathematical framework used to motivate the incorporation of these two ideas.

2. Operations Research

Generally speaking, O.R is a branch of math that seeks to make a best decision, typically through an optimization model. In the real world some of the objectives to optimize are often concepts such as cost, profit, distance, loss, risk, and value. Two examples of problems in the O.R domain are the Bin packing problem, and the facilities location problem.

This type of model is extremely valuable to modern syndicates as the premise of minimizing *extraneous* costs (payroll hours, health insurance, paid time off, etc.), while maximizing profit is at the core of these syndicates existence. Herein lies the moral motivation for the necessity of the merger of equity and O.R. Mathematics is defined by its dispassionate and objective nature. This means that factories, warehouses, boardrooms, and bad actors use O.R to optimize profit above all others. Conspicuously absent from these calculations is the human element. In the eyes of the corporate machine profit reigns supreme, and all other factors remain unseen. I find it is my prerogative to rage against such machines. Thus, it is necessary to incorporate people into these calculations, into these metrics, into these models. People must be optimized, we must be taken into account.

This paper focuses on the Bin Packing problem (B.P.P) which is a well studied O.R problem with many applications such as; load balancing and memory allocation in parallel processing [Ekici, 2022] 3D bin packing and load balancing [Erbayrak et al., 2021], and the aircraft maintenance task problem.[Witteman et al., 2021]. Note that only two general models for the B.P.P will be presented and analyzed, rather than focusing on any specific problem. This generality is deliberate in order to ensure the accessibly and relevance to all parties interested in such a formulation. Such a general model however, is both abstract and hypothetical. These traits make conceptualizing the merger of ideas difficult. To remedy this, we present a specific example.

2.0.1. *Existing Research.*

The Facilities Location problem is another well studied O.R. problem. To motivate it, consider the following example: Imagine you are a city planner, and want to add two grocery stores inside city limits. Where do you place them? The classic solution to this problem is typically an optimization integer programming model, which minimizes the average distance each resident must travel to reach one of the stores. In other words every resident has, ostensibly, equal access to these stores.

This solution only considers the *average* distance of residents to store, and thus may produce a solution wherein the majority of residents have a small travel time, but some residents are forced to travel a significant distance in order to reach the store.

This classic solution is not equitable. An equitable solution would be one that mitigates or minimizes the distance the outlier residents need to travel. This example is used to illustrate an intrinsic inequality in the classic solution. Despite the average being minimized, it is important to remember there are people who are forced to endure an unequal share of the burden. For more information on this specific problem, and the what an equitable distribution of locations looks like, see [Horton, 2021].

2.0.2. *The Bin Packing Problem (B.P.P).*
O.R and equity can successfully be merged, as is evidenced by [Horton, 2021] and [Logan et al., 2021]. We focus now on extending the integration of equity into another O.R problem, the bin packing problem. We consider a general model in order to ensure the accessibility and relevance to all parties interested in such a formulation. The below model is considered a classic formulation of the bin packing problem. In it, we are given a set of items, each of their own size, and a set of bins, with their own volume. The goal of is to pack all items in as few bins as possible, while respecting some logical structure. i.e. the total size of all items in any bin must not exceed the bins volume. Additionally, we must pack every item in some bin.

Due to the structure of this classic model, we cannot easily incorporate equity. To remedy this, we transform the classic model into another, equivalent, model. We refer to this new model as the Syndicate model. From there we'll derive another, similar, model which incorporates a metric of equity called the Kolm-Pollak Equally-Distributed Equivalent. The final analysis of data, and the conclusion of this paper will rely on juxtaposing the Syndicate model and the Equity model. We now transform the Classic model by incrementally including information. To begin, Suppose there are multiple syndicates from which bins may be purchased. We include a new set of syndicate:

S $= \{1, \ldots, s, \ldots, n\}$ - Set of Syndicates selling bins, from 1 to n using index s

Corresponding to this new set is a new parameter, which indicates a particular bin was purchased from a particular syndicate:

$$A_{b,s} = \begin{cases} 1 & \text{if bin b is purchased from syndicate s} \\ 0 & \text{otherwise} \end{cases}, \forall b \in B \, \forall s \in S$$

Finally, observe how the objective function from the Classic method can be modified to include the price of the bin, and whether or not it was purchased from a given syndicate.

$$\min \quad \sum_{s \in S} \sum_{b \in B} A_{b,s} \cdot P_b \cdot X_b, \forall s \in S$$

For simplicity sake, we assign the second sum its own variable:

$$D_s = \sum_{b \in B} A_{b,s} \cdot P_b \cdot X_b, \forall s \in S$$

Thus we have the Syndicate model

Sets

B = $\{1, \ldots, b \ldots, m\}$ - Set of bins to fill, from 1 to m using index b

S = $\{1, \ldots, s, \ldots, n\}$ - Set of Syndicates selling bins, from 1 to n using index s

I = $\{1, \ldots, i, \ldots p\}$ - Set of items to pack, from 1 to p using index i

Parameters

C_b - Capacity of bin b, $\forall b \in B$

P_b - Price to purchase bin b, $\forall b \in B$

$A_{b,s} = \begin{cases} 1 & \text{if bin b is available from syndicate s} \\ 0 & \text{otherwise} \end{cases}, \forall b \in B \, \forall s \in S$

E_i - Size of item i, $\forall i \in I$

Decision Variables

$X_b = \begin{cases} 1 & \text{if bin b is packed with any item} \\ 0 & \text{otherwise} \end{cases}, \forall b \in B$

$Y_{i,b} = \begin{cases} 1 & \text{if item i is packed into bin b} \\ 0 & \text{otherwise} \end{cases}, \forall i \in I, \forall b \in B$

D_s - Total money paid to syndicate s (for bins), $\forall s \in S$

Model

$$\min \quad \sum_{s \in S} D_s \tag{2.1}$$

$$\text{s.t} \quad \sum_{b \in B} Y_{i,b} = 1, \forall i \in I \tag{2.2}$$

$$\sum_{i \in I} E_i \cdot Y_{i,b} \leq C_b \cdot X_b, \forall b \in B \tag{2.3}$$

$$D_s = \sum_{b \in B} A_{b,s} \cdot P_b \cdot X_b, \forall s \in S \tag{2.4}$$

$$X_b \in \{0, 1\} \; \forall b \in B \tag{2.5}$$

$$Y_{i,b} \in \{0, 1\} \; \forall i \in I, \; \forall b \in B \tag{2.6}$$

Reference

2.1 Objective is to minimize the total amount paid, summed over all syndicates.

2.2 Item may only be packed into one bin.

2.3 Total size of items $(i_1, i_2, \dots)$ in bin b must be be less than or equal to capacity of bin b.

2.4 Dark money paid to syndicate s equals sum of the bins purchased (from the syndicate) times the price of the bins.

2.5 Binary choice, is the bin packed or no?

2.6 Binary choice, is the bin packed with the specific item or no?

The solution to this problem is the ‘best’ distribution of bins, which much be purchased from specific syndicates. In this current version, the solution will be a distribution of bins of sufficient volume which are purchased from the syndicate who sells them for the least amount of money. In other words, all bins are purchased for the lowest cost, regardless of which syndicate sells them.

When we introduce equity, we will change the objective function and thus the goal of the model. The new solution will be a distribution of bins of sufficient volume which are purchased from syndicates in an *equitable* manner. As we are working with a dis-amenity, the solution to the equity model will be a set of bins which are purchased from the fewest number of syndicates possible. We next motivate equity further, and provide the specifics on the mathematics of equity.

3. Equity

We motivate equity with an example involving a distribution of resources, hypothetical populations, and scarcity: Consider a population, a necessary and finite resource for this population, and a set of corresponding sub-populations. Equity asserts that a morally correct distribution of that resource may involve a smaller sub-population receiving more of the resource than another, larger, sub-population. To illustrate this, consider the visual metaphor below.

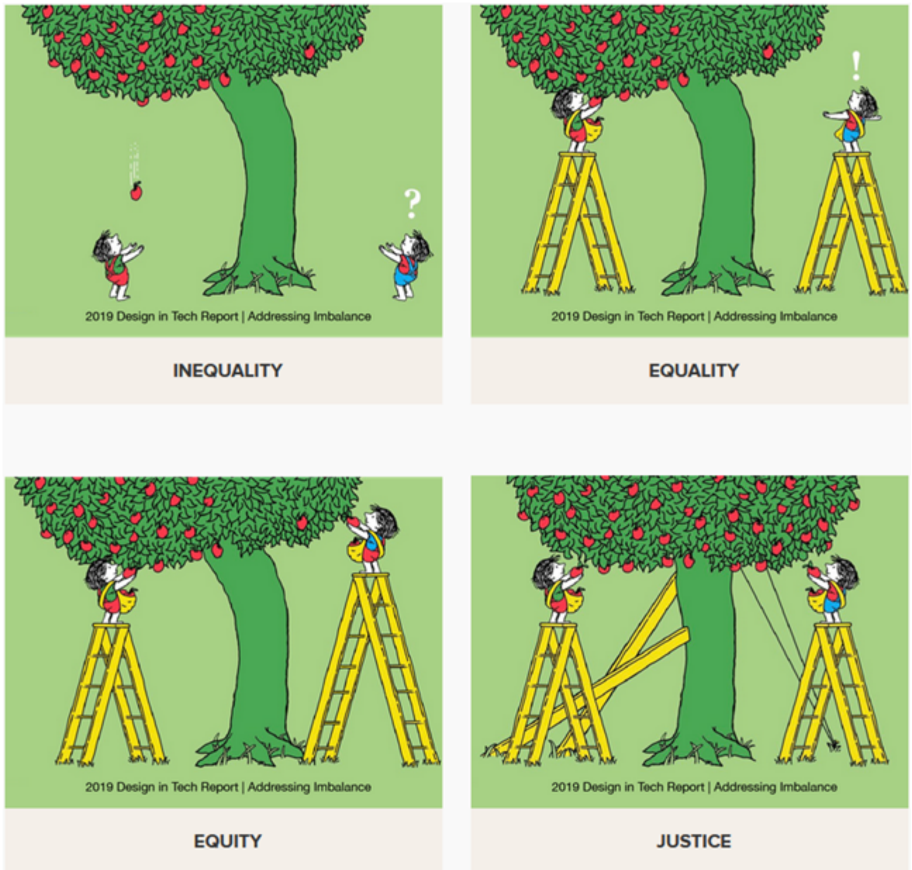

FIGURE 1. Equity

[Ruth, 2019]

This paper assumes the necessity of equity in the modern world, and will present and analyze the two optimization models discussed before. This analysis will be on six data points, 3 for each model; time spent computing the solution, the total amount spent on bins, and the inequality index, or the measure of equity in the solution.

3.1. **Inequality Measures.** As was mentioned in the introduction, measuring inequality in a system began as a purely economic exercise. From there equity has been measured in the context of environmental justice [Maguire and Sheriff, 2011] [Sheriff and Maguire, 2020], and in operations research via the facilities location problem. [Horton, 2021]. This paper extends the incorporation of equity in O.R. by focusing on equity in the bin packing problem.

In order to numerically motivate equity, we discuss inequality indices, and equally distributed equivalents. An inequality index is a broad term for a set of functions which take as an input a relevant distribution, and outputs a number. This distribution may be of amenities (something people generally want more of) or of dis-amenities. The resulting number is used to rank distributions. The inequality indexes presented below are robust and historically significant, but insufficient for our purposes. [Sheriff and Maguire, 2020] explains: "Although inequality indexes are derived from individual preferences, it is important to recognize that the index values are not normative rankings themselves (Kaplow, 2005). That is, just because a distribution is less equal than another, does not necessarily mean that it is less desirable; total levels are also relevant. Consequently, we advocate use of equally distributed equivalents (EDEs) for evaluating distributions. EDEs are derived from the same preference structure as inequality indexes, but take both levels and dispersion into account. The EDE answers the question "What is the level of risk that would make an individual indifferent between a distribution in which everyone received that risk and the actual unequal risk distribution?""

There are many types of inequality indexes available to use in measuring a distribution. This diversity of options forces us to consider what an 'ideal' measure should look like. From the existing literature [Adger, 2000], [Blackwood and Lynch, 1994],[Logan et al., 2021], [Sheriff and Maguire, 2020] . there are several properties which our chosen measure should satisfy:

- <u>Symmetry</u>: Inequality of a population is based solely on the distribution of the quantity in question and no other rankings.
- <u>Population independence</u>: The number of individuals does not influence the measure of inequality.
- <u>Scale dependence</u> : The inequality measure should reflect the total sum of the quantity measured. For instance, high exposure to burdens (for example hazards or health risks) would be penalized.
- <u>Principle of transfers</u>: If a quantity is redistributed from an advantaged individual to a disadvantaged individual, the inequality should decrease.
- <u>Satisfies the mirror property</u>: The measure can be used for distributions of both amenities (desirable) and dis-amenities (undesirable) quantities.
- <u>Separable</u>: The measure can be used to examine the inequality between subgroups (e.g., different demographic groups), and therefore can incorporate consideration of need or vulnerability, and, subsequently, inequity.
- <u>Multivariate</u>: The measure can be used to evaluate multiple quantities and the correlation of advantages or disadvantages.

At this point there is no measure which satisfies all these criteria. This leads to a fundamental question. Which should one use to measure a distribution in which circumstances? This is not an easy question to answer, and debate rages on which to use. We now give a brief overview of common inequality indexes.

3.1.1. *Gini Index.* Originally presented by Corrado Gini in his 1912 book Variabilità e Mutabilità" (Variability and Mutability), this function has (at least) 13 different formulations. In its current iteration, it is a monotionically increasing function which represents the area between the Lorenz curve, and the actual distribution. While simple, this Index has some notable intrinsic deficiencies when measuring inequality. To motivate this phenomenon, consider the Gini Index's measure of a distribution of dis-amenities. As the name suggests, a dis-amenity is an item or concept we would want less of. Pollution, economic hardship, and disease are all examples of bads.

When ranking two distributions of bads, the Gini Index would prefer a distribution which makes everyone equally as miserable. To illustrate this, consider the financial collapse of 2007, the most extensive and destructive financial crisis since the great depression. [Gopinath, 2020]
Consider a classic bin packing formulation on a simplified version of the problem, where we're using the Gini index to measure the distribution of units of economic hardship (items) into the 300 million people living in the united states at that time (bins). Consider two distinct distribution of that economic hardship;

- Distribution 1: Economic hardship units (EHU) are distributed equally among the population, affecting all citizens roughly equally, resulting in an increase of homelessness in american families by 30% [Foundation, 2012]
- Distribution 2: EHU's are are only distributed to a specific group, minimizing the effect on the overall population.

According to the Gini Index, the first distribution is preferable. Contrary to this, however, is the knowledge of situations in which a less equal distribution of bads is preferable to the population as a whole. In the previous example, the question of which group should bear the brunt of bads is obvious. In more general examples the answer is not so clear. This is further evidence that equity in mathematical formulation of problems is necessary for the success and health of the population as a whole.

3.1.2. *Atkinson measure.* Originally created as a purely economic measure, Atkinson [Atkinson, 1970] formulated a method to measure inequality in the distribution of income. This method, while useful in some contexts is unsuitable for our purposes, as in it's default form, it is incapable of providing a reasonable solution when measuring a distribution of dis-amenities. In other words, this measure does not satisfy the mirror property.

The failure of the Atkinson measure to satisfy the mirror property immediately disqualifies from this paper as a reasonable metric. To reiterate, this paper is interested in developing a *general* formulation of the bin packing problem, in order to remain relevant to a large audience. Thus, our choice of measure must be able to measure a distribution of dis-amenities. The necessity of this condition is obvious, as both our models involve minimizing cost, intrinsically a dis-amenity.

3.1.3. *Kolm-Pollak.* Finally, we come to the last method, and the focus of the paper. The Kolm-Pollak EDE is a robust type of inequality index. In general, this EDE is preferable to the other inequality indices listed in the previous section, as it satisfies the majority of the ideal properties. It does not, however, satisfy all of them. Its general formulation for a distribution X with sample size N is:

$$\Xi(X) = -\frac{1}{\kappa}\ln\left(\frac{1}{N}\sum_{i\in N} e^{-\kappa X_i}\right)$$

Observe κ, this is a subjective constant which we define to be aversion to inequality in the population. This parameter allows the KP-EDE to produce a reasonable measure when measuring both amenities and dis-amenities. In fact, κ is a normalized value of the inequality aversion parameter ϵ proposed by [Atkinson, 1970]. We motivate ϵ, in order to provide context to κ.

When measuring a distribution of amenities ϵ is positive, and the corresponding KP-EDE will penalize large values in that distribution. We present an illustrative example in an economic context: Given two distributions, one where the majority of the populace has a similar level of wealth, and another where fewer people have more wealth, the KP-EDE will rank the former higher.

When measuring a distribution of dis-amenities ϵ is negative, and the corresponding KP-EDE will penalize small values in that distribution. We present an illustrative example in a environmental justice context: Given two distributions, one where the majority of the populace are exposed to similar levels of pollution, and another were fewer people have more pollution, the KP-EDE will rank the latter higher.

For the equity model, we are considering a distribution of costs (dis-amenities) so we choose ϵ = -1. The resulting κ value is nothing more than a normalized version of ϵ. As before, κ is a subjective inequality aversion parameter. We can normalize κ as follows:

$$\kappa = \frac{\sum_{i=1}^{N} z_i}{\sum_{i=1}^{N}(z_i)^2}\epsilon$$

where z_i is the i^{th} value of the distribution in question, and N is the number of values in the distribution. Using notation above we now define the KP-EDE for our problem as

$$\Xi(D_s) = -\frac{1}{\kappa}\ln\left(\frac{1}{|S|}\sum_{s\in|S|} e^{-\kappa D_s}\right)$$

Where $|S|$ is the number of syndicates and D_s is the total paid to each syndicate.

4. Incorporating Equity

At last we come to the mechanics of using our measure of equity, the Kolm-Pollak EDE, in the bin packing problem. Recall the formulation of the KP-EDE.

$$-\frac{1}{\kappa}\ln\left(\frac{1}{|S|}\sum_{s\in|S|}e^{-\kappa D_s}\right)$$

Due to computational limitations, we cannot simply substitute the KP-EDE into a linear program's objective function. To that end, we use common techniques to simplify the KP-EDE into a form we can solve.

As previously discussed, the scalar multiple $\frac{-1}{\kappa}$ can be discarded from the objective function as it is a scalar multiple, and thus does not affect the solution to the optimization problem.

Thus the KP-EDE is simplified to

$$\ln\left(\sum_{k\in T}e^{-\kappa E_k}\right)$$

Recall that the natural logarithm function is monotonically increasing which means it achieves its maximum value at the same points as operand. Thus we can discard the natural logarithm, and focus on the summation

$$\sum_{k\in T}e^{-\kappa E_k}$$

This leads us to the Equity model, presented on the next page.

Sets

B = $\{1,\ldots,b\ldots,m\}$ - Set of bins to fill, from 1 to m using index b

S = $\{1,\ldots,s,\ldots,n\}$ - Set of Syndicates selling bins, from 1 to n using index c

I = $\{1,\ldots,i,\ldots p\}$ - Set of items to pack, from 1 to p using index i

Parameters

C_b - Capacity of bin b, $\forall b \in B$

P_b - Price to purchase bin b, $\forall b \in B$

E_i - Size of item i, $\forall i \in I$

D_s - Total money paid to syndicate s (for bins), $\forall s \in S$

$$A_{b,s} = \begin{cases} 1 & \text{if bin b is purchased from syndicate s} \\ 0 & \text{otherwise} \end{cases}, \forall b \in B\, \forall s \in S$$

Decision Variables

$$X_b = \begin{cases} 1 & \text{if bin b is packed with any item} \\ 0 & \text{otherwise} \end{cases}, \forall b \in B$$

$$Y_{i,b} = \begin{cases} 1 & \text{if item i is packed into bin b} \\ 0 & \text{otherwise} \end{cases}, \forall i \in I, \forall b \in B$$

Model

$$\min \quad \sum_{s\in S} e^{-\kappa D_s} \tag{4.1}$$

$$\text{s.t} \quad \sum_{b\in B} Y_{i,b} = 1, \forall i \in I \tag{4.2}$$

$$\sum_{i\in I} E_i \cdot Y_{i,b} \le C_b \cdot X_b, \forall b \in B \tag{4.3}$$

$$D_s = \sum_{b\in B} A_{b,s} \cdot P_b \cdot X_b, \forall s \in S \tag{4.4}$$

$$X_b \in \{0,1\}\, \forall b \in B \tag{4.5}$$

$$Y_{i,b} \in \{0,1\}\, \forall i \in I,\, \forall b \in B \tag{4.6}$$

Reference

4.1 Objective is to minimize the total amount paid, summed over all syndicates.

4.2 Item may only be packed into one bin.

4.3 Total size of items $(i_1, i_2, \ldots)$ in bin b must be be less than or equal to capacity of bin b.

4.4 Dark money paid to syndicate s equals sum of the bins purchased (from the syndicate) times the price of the bins.

4.5 Binary choice, is the bin packed or no?

4.6 Binary choice, is the bin packed with the specific item or no?

5. Methods and Data

We now turn our focus to the mechanics of the project, and present in-depth details on both models (Syndicate and Equity), the solver, the data, the results, and everything in-between.

5.1. **Models.**

As was discussed previously, the Syndicate model Dr.Speakman and myself developed is a linear integer programming optimization model. This model is equivalent to the classic bin packing model, and was developed in order to simplify the comparison of results from a model where the KP-EDE was not used, to one where it was used.

When we substitute the KP-EDE into the objective function the Syndicate model becomes the equity model. Due to the formulation of the KP-EDE, the corresponding equity model is non-linear. As a result, computation times increase drastically. It is possible there is an equivalent linear formulation of this model, but I was not able to derive it. This linearization of the equity model is an area where I would continue to research this project, given more time.

Technically it would be possible to compare the results of the classic bin packing model to the equity model but as one might expect, the mechanics of tracking the difference in parameters and objectives from these two would detract from the main focus of this paper. Namely, what is the relationship between the increased cost vs the increase in equity in the two models. In other words, does the increase in the equity offset the increase in cost?

I have coded both the Syndicate and Equity models in Python using the SCIP optimization tools via PySCIPOpt. This choice of solver was deliberate as SCIP is both a non-linear solver, and is licensed under the open source Apache 2.0 license and is "software for the public good". [Apache, 2023]

5.2. **Data Creation.** Due to the unexpected complexity in coding coding the models, I was not able to extract real-world data corresponding to this problem, and thus all the data I used is random. As the data used is random, I relied on a deliberate form of data control. Notably, each component of the data has a specific column and a specific name across all of the dataframes and functions. I anticipated it would be easier to ensure data integrity by preforming calculations on simple numbers in specific columns, rather than attempt to extract, analyze and preform calculations on non-scrubbed or specially formatted data.

In a future section I will illustrate how all the python files fit together, but for now I will explain the details of the data creator file.

This python file takes 3 integers as input, corresponding to the number of bins, the number of items, and the number of syndicates for a particular simulation. From there the item sizes, and bin information is populated using numpy and the randomint() function. In addition, there are lists, dataframes and values set up for future use.

```
def create_data(num_syndicates, num_items, num_bins):


item_sizes = pd.Series(np.random.randint(2,5) for i in range(num_items))
bin_vols = pd.Series(np.random.randint(5,9) for b in range(num_bins))
price_of_bins = [np.random.randint(3,10) for b in range(num_bins)]


input_matrix = []
a_matrix_col_s = []
bin_price = []
bin_volume = []
bin_data = pd.DataFrame({})
unique_bins = 0
```

⋮

From there, we loop through the number of syndicates in order to get the availability matrix, which is the parameter $F_{s,b} \in \{1, 0\}$ from the syndicate and equity model representing whether the syndicate has the bin available for purchase or not.

⋮

```
for s in range(num_syndicates):
    #create random list of 1's/0's
    a_matrix_col_s = [np.random.randint(0,2) for i in range(num_bins)]

    ⋮
```

As you can see, this is also randomly generated. In this same loop we calculate the total volume of bins offered by the syndicate by using a simple sum-product calculation on the availability matrix, and the bin volume. We must ensure the random availability matrix yields a total volume of at least $\lceil \sum_{i \in I} S_i / \text{number of syndicates} \rceil$, in order to ensure enough volume is generated to contain all items.

⋮

```
    #Ensure sufficient bin volume per syndicate
    sumProduct_LowerBound = math.ceil(sum(item_sizes/num_syndicates))

    #sum_{num_synd} bin_volume(b) * a_matrix_col_s(b) for all bins
    sumProduct_bincap_x_pmatrix = sum(map(operator.mul, bin_vols, a_matrix_col_s))

    #What if the input_matrix did not yield enough bin volume?
    #We randomly change elements of input_matrix into 1, to increase the volume
    while(sumProduct_bincap_x_pmatrix < sumProduct_LowerBound):
        a_matrix_col_s[np.random.randint(0,num_bins)] =1
        sumProduct_bincap_x_pmatrix = sum(map(operator.mul, bin_vols, a_matrix_col_s))

     ⋮
```

One of the flaws of this program is that some double work is required to accommodate the syndicate and equity models. Both of those assume that a bin available from two different syndicates is actually two unique bins. In its current iteration this file is designed to generate information (volume, cost, availability) for a fixed number of bins. This means the file cannot determine the number of unique bins until it has randomly created the entire availability matrix. Once it has done that, it calculates the number of unique bins by determining where (and thus how many) the 1's are in the availability matrix per each syndicate.

```
     ⋮

#We need to ensure that each duplicate from the input_matrix is treated as a uniquie item
#thus we redo some work, and create the new a_matrix
a_matrix_col_s = [0]*unique_bins

location_of_ones_in_a_matrix_col_s = np.where(a_matrix_col_s)[0]
number_of_ones_in_a_matrix_col_s = len(location_of_ones_in_a_matrix_col_s)

     ⋮
```

From there it becomes a bookkeeping matter, appending lists with relevant information, and returning the dataframe object.

```
     ⋮

for b in range(number_of_ones_in_a_matrix_col_s):
    a_matrix_col_s.append(1)
```

```
    bin_price.append(price_of_bins[location_of_ones_in_a_matrix_col_s[b]])
    bin_volume.append(bin_vols[location_of_ones_in_a_matrix_col_s[b]])

df_temp = pd.DataFrame({f"available_from_syndicate_{s}?":pd.Series(a_matrix_col_s)})

bin_data =pd.concat([bin_data,df_temp],axis = 1).fillna(0)

unique_bins += len(np.where(a_matrix_col_s)[0])

availabilty_matrix.append(a_matrix_col_s)

df_temp = pd.DataFrame({
"Syndicate_Names": pd.Series(i for i in range(num_syndicates)),
"Item_Size": pd.Series(np.random.randint(2,5) for i in range(num_items)),
"Bin_Volume":pd.Series(bin_volume),
"Price_to_purchase_bin":pd.Series(bin_price)
 })

bin_data = pd.concat([bin_data,df_temp],axis = 1)

#need to re-arrange to meet the criteria of the original data creator
cols = bin_data.columns.tolist()
cols = cols[-1:] + cols[:-1]
cols = cols[-1:] + cols[:-1]
cols = cols[-1:] + cols[:-1]
cols = cols[-1:] + cols[:-1]
bin_data = bin_data[cols]
return(bin_data)
```

5.3. **Kappa Creation.** As discussed, κ is a normalized value of the inequality aversion parameter, ϵ. For this problem we're measuring a distribution of dis-amenities, so we choose $\epsilon = -1$. Note that we need to scale ϵ in order to provide a reasonable answer for our distribution. To that end, we must calculate a 'reasonable' κ value. We do this by calculating the cost if each item is packed into it's own bin. This is reasonable, as each item necessarily needs to be packed in a bin, and all the data used is random. As such, there is no natural or observable way to create a more equitable distribution of items in bins. The code below shows exactly how this process is preformed.

```
def kappa(data):
    total_cost_of_purchased_bins = 0
    for j in range(len(data["Item Size"])):
        bin_counter = 0
        while data["Item Size"].iloc[j] >= data["Bin Volume"].iloc[bin_counter]:
            bin_counter += 1
        total_cost_of_purchased_bins += data["Price to purchase bin"].iloc[bin_counter]
    return(-len(data["Syndicate Names"])/total_cost_of_purchased_bins)
```

5.4. **Cluster.**

As we will motivate in the next section, the time to compute these programs increases drastically with small size increases. To combat this, we utilize the University of Colorado Denver CCM Clusters, specifically the Alderaan cluster [Mandel, 2023]. These clusters are generously funded by the National Science Foundation OAC-2019089. [NSF, 2023], and tirelessly supported by Dr. Jan Mandel. [Google, 2023]

These clusters allow for complex computations to be performed continuously, up to a seven-day limit. Lastly, I present a graphic to motivate to code execution structure:

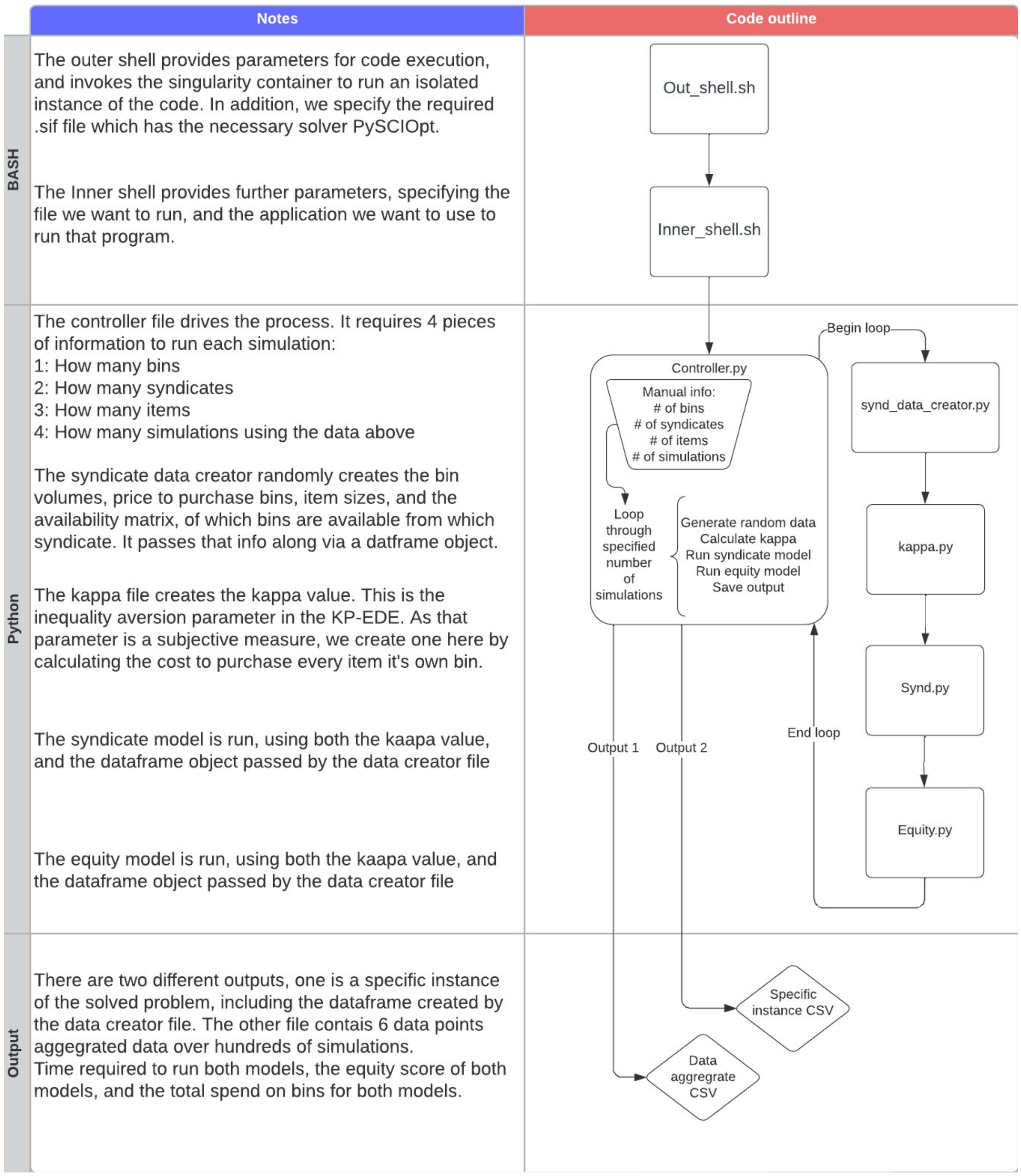

FIGURE 2. Code Execution flow

The two output files will be motivated in the next section.

6. Results & Discussion

We now present the results of the analysis. As mentioned previously, the analysis will take place on six data points, 3 for each model; time spent computing the solution, the total amount spent on bins, and the inequality index, or the measure of equity in the solution.

6.1. **Output files.** As mentioned before, the analysis relies on two output files. One of these files is a specific instance of a simulation for both models, the other is an aggregate collection of all simulations.

The below picture is an individual simulation of the syndicate and equity model for size [5,5,5]. Column A is the number of unique bins generated by the availability matrix. Columns F-J contain the indicator of whether that bin is available from that particular syndicate. You'll notice this is well structured, which was designed in the data creator file. Column K corresponds to col A, and indicates if that bin is filled. Column L corresponds to column C, and represents which bin each item is packed into. Columns M-N, O-P, Q-R, S-T, and U-V all have the same layout. The first entry, Col M for example, correspond again to col A. This indicates if any of those bins are purchased from the specific syndicate. The second entry, Col N, is a list representation of that same information. In the case no bin is purchased, the list is empty.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Uniquie Bins	Syndicate Names	Item Size	Bin Volume	Price to purchase bin	available from syndicate 0?	Avail 1?	Avail 2?	Avail 3?	Avail 4?	bins filled	items packed in below bin	purchased from syndicate 0?	bins purchased from syndicate list	B.P 1?	L 1	B.P 2?	L 2	B.P 3?	L 3	B.P 4?	L 4
0	0	3	6	8	1	0	0	0	0	0	3	0	[]	0	[3]	0	[]	0	[9]	0	[]
1	1	3	8	6	0	1	0	0	0	0	9	0		0		0		0		0	
2	2	2	8	7	0	1	0	0	0	0	9	0		0		0		0		0	
3	3	4	8	4	0	1	0	0	0	1	3	0		1		0		0		0	
4	4	3	6	9	0	1	0	0	0	0	9	0		0		0		0		0	
5			8	6	0	0	1	0	0	0		0		0		0		0		0	
6			8	7	0	0	1	0	0	0		0		0		0		0		0	
7			6	9	0	0	1	0	0	0		0		0		0		0		0	
8			8	6	0	0	0	1	0	0		0		0		0		0		0	
9			8	4	0	0	0	1	0	1		0		0		0		1		0	
10			6	9	0	0	0	1	0	0		0		0		0		0		0	
11			8	6	0	0	0	0	1	0		0		0		0		0		0	
12			8	7	0	0	0	0	1	0		0		0		0		0		0	
13			8	4	0	0	0	0	1	0		0		0		0		0		0	

Figure 3. Syndicate results, [5,5,5]

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Uniquie Bins	Syndicate Names	Item Size	Bin Volume	Price to purchase bin	available from syndicate 0?	Avail 1?	Avail 2?	Avail 3?	Avail 4?	bins filled	items packed in below bin	purchased from syndicate 0?	bins purchased from syndicate list	B.P 1?	L 1	B.P 2?	L 2	B.P 3?	L 3	B.P 4?	L 4
0	0	3	6	8	1	0	0	0	0	0	3	0	[]	0	[3]	0	[]	0	[9]	0	[]
1	1	3	8	6	0	1	0	0	0	0	3	0		0		0		0		0	
2	2	2	8	7	0	1	0	0	0	0	3	0		0		0		0		0	
3	3	4	8	4	0	1	0	0	0	1	9	0		1		0		0		0	
4	4	3	6	9	0	1	0	0	0	0	9	0		0		0		0		0	
5			8	6	0	0	1	0	0	0		0		0		0		0		0	
6			8	7	0	0	1	0	0	0		0		0		0		0		0	
7			6	9	0	0	1	0	0	0		0		0		0		0		0	
8			8	6	0	0	0	1	0	0		0		0		0		0		0	
9			8	4	0	0	0	1	0	1		0		0		0		1		0	
10			6	9	0	0	0	1	0	0		0		0		0		0		0	
11			8	6	0	0	0	0	1	0		0		0		0		0		0	
12			8	7	0	0	0	0	1	0		0		0		0		0		0	
13			8	4	0	0	0	0	1	0		0		0		0		0		0	

Figure 4. Equity results, [5,5,5]

Unfortunately the results I chose to output do not illustrate the 'best case' differences between the equity model and the syndicate model. As we'll see in the next section, for size [5,5,5] the largest differences were generated by simulation 7 and 85. I attempted to limit computation time by outputting results every 10 simulations.

The second output file represents a more aggregate picture, and contains information not stored on the individual files. Simulation 7 is highlighted showing one of the largest differences in a simulation of this size. Unfortunately, I did not output the same information from the first file for this simulation.

Simulation Number	Simulation size	Run time - syndicate model	Run time - equity model	Total spent - syndicate model	Total spent - equity model	Inequality index, syndicate model	Inequality index, equity model
0	[5, 5, 5]	0.034822941	0.083558798	8	8	6.297442541	6.297442532
1	[5, 5, 5]	0.016434908	0.071810722	14	14	8.436563657	8.436562657
2	[5, 5, 5]	0.014610767	0.03406024	6	6	5.909982829	5.909982829
3	[5, 5, 5]	0.02485919	0.220913649	15	16	13.19116633	11.87312731
4	[5, 5, 5]	0.020640373	0.032467365	9	9	10.15484549	10.15484549
5	[5, 5, 5]	0.016018629	0.02188015	12	12	7.140252037	7.140252037
6	[5, 5, 5]	0.01821208	0.634047985	8	9	8.607342974	8.35100005
7	[5, 5, 5]	0.021677017	0.078855038	11	12	20.11019792	11.87312731
8	[5, 5, 5]	0.013413191	0.04517436	6	6	6.070126019	6.070125157
9	[5, 5, 5]	0.017848969	0.055671692	6	6	8.436563657	8.436563657
10	[5, 5, 5]	0.017596483	0.04254961	6	6	8.436563657	8.436563656
11	[5, 5, 5]	0.019583464	0.236427784	9	9	10.15484549	10.15484549
12	[5, 5, 5]	0.014148712	0.02036047	9	9	6.186837275	6.186837275
13	[5, 5, 5]	0.019147873	0.137847424	10	10	7.127213395	7.127212491
14	[5, 5, 5]	0.018153667	0.127254486	13	13	7.165688499	7.165688499
15	[5, 5, 5]	0.015171051	0.280331612	12	12	9.640233845	9.640232365
16	[5, 5, 5]	0.015688181	0.047662735	10	10	6.5587041	6.558704092
17	[5, 5, 5]	0.015018463	0.057209969	19	21	9.673358551	9.304809614
18	[5, 5, 5]	0.015461683	0.033599854	6	6	7.234000033	7.234000033
19	[5, 5, 5]	0.015831232	0.035654783	8	8	6.297442541	6.297442476
20	[5, 5, 5]	0.017363548	0.751072884	9	9	10.15484549	10.15484548
21	[5, 5, 5]	0.017481804	0.029406071	6	6	5.79122485	5.79122485
22	[5, 5, 5]	0.035902023	0.22780323	12	12	6.895468082	6.738958891
23	[5, 5, 5]	0.016845226	0.057211161	15	15	9.779131612	8.243605385
24	[5, 5, 5]	0.01522994	0.105311155	12	12	8.436563657	8.436562657
25	[5, 5, 5]	0.0157938	0.047784567	10	10	8.436563657	8.436562657
26	[5, 5, 5]	0.020229101	0.105952024	8	8	8.436563657	8.436563656
27	[5, 5, 5]	0.018970013	0.111031532	12	12	7.312384857	7.312384847
28	[5, 5, 5]	0.01529932	0.049325466	8	8	8.607342974	8.607342974
29	[5, 5, 5]	0.019516706	0.093650579	6	6	6.297442541	6.297442154
30	[5, 5, 5]	0.017335415	0.298178434	8	8	8.436563657	8.436563655
31	[5, 5, 5]	0.019647121	0.060457468	14	14	7.797750588	7.797750588
32	[5, 5, 5]	0.014746428	0.041415453	12	12	13.10733793	10.15484549
33	[5, 5, 5]	0.013033152	0.153309345	6	6	7.234000033	7.234000033
34	[5, 5, 5]	0.015712023	0.027919292	11	11	9.280740895	9.280740895
35	[5, 5, 5]	0.017289639	0.157274485	14	14	9.511949723	9.511949723

FIGURE 5. Aggregate results, [5,5,5]

6.2. **Simulation run times.** We next discuss the time taken to complete the calculations, in seconds, for both models. As as reminder, the data creator file, and thus the models, are solved based on three inputs; the number of bins, the number of items, and the number of syndicates. These three integers are presented a tuple going forward, of form [B,I,S], and dictate the size of the simulations.

The elapsed time for each simulation size is presented below. Note: the following time values have been excluded from 7 and 8 respectively, in order to increase readability of the graph

$[10, 10, 10]$: 6035, 1008, 918, 400, 91, 66, 61

$[15, 10, 10]$: 18731, 14186, 9773, 5807, 5169, 1702, 1434

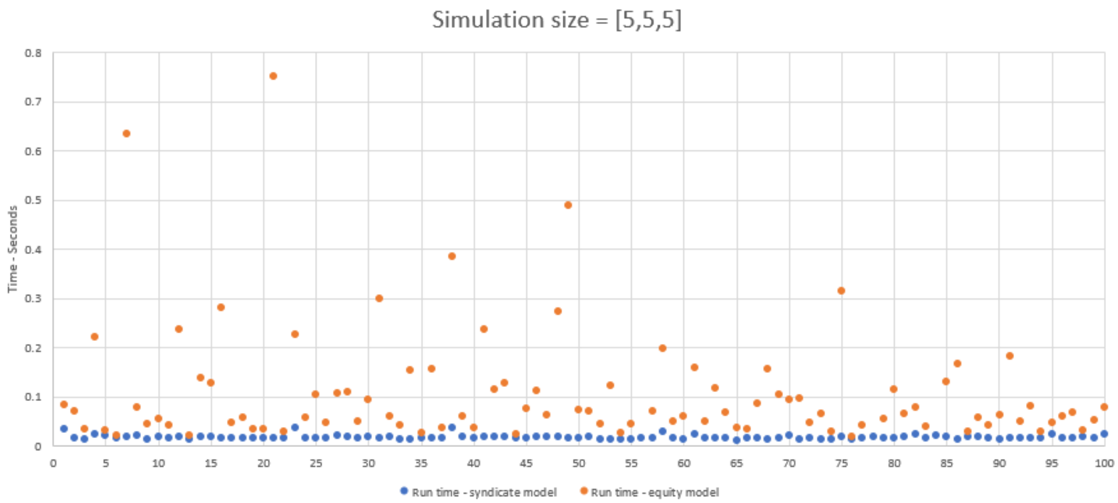

FIGURE 6. Time [5,5,5]

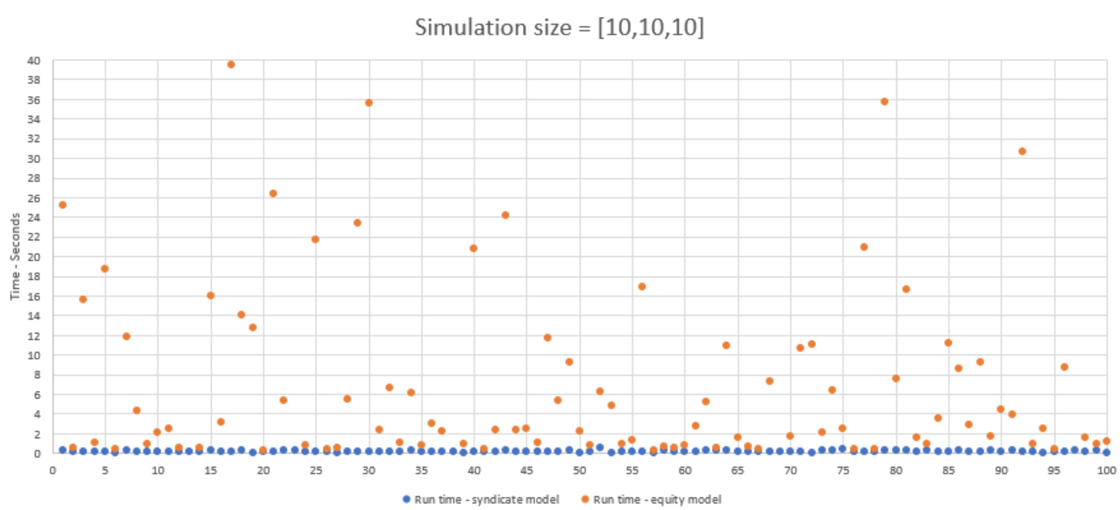

FIGURE 7. Time [10,10,10]

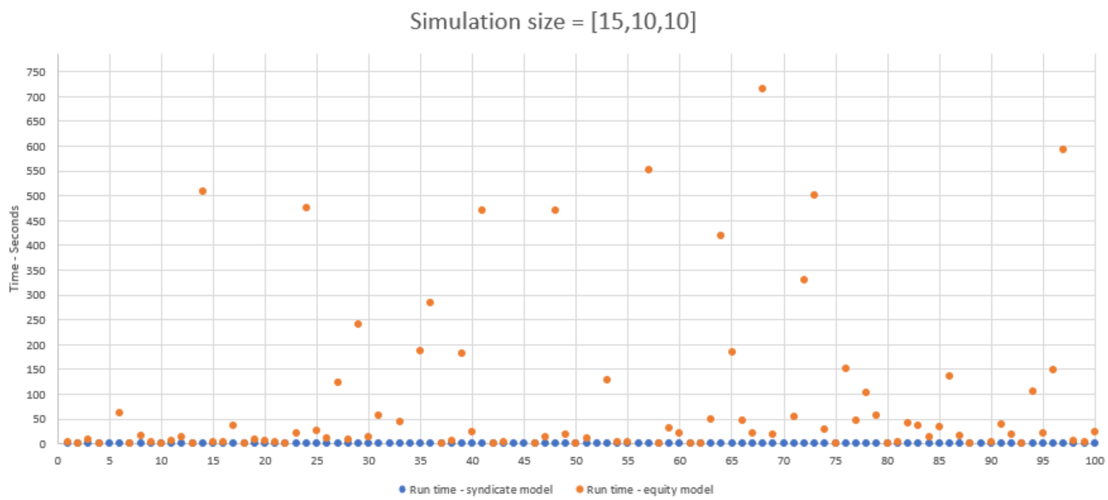

FIGURE 8. Time [15,10,10]

Although the graphs above are illustrative of part of the story, in order to fully motivate the vast differences in time, we present one more graph. The graph below contains the times of all simulations for both models.

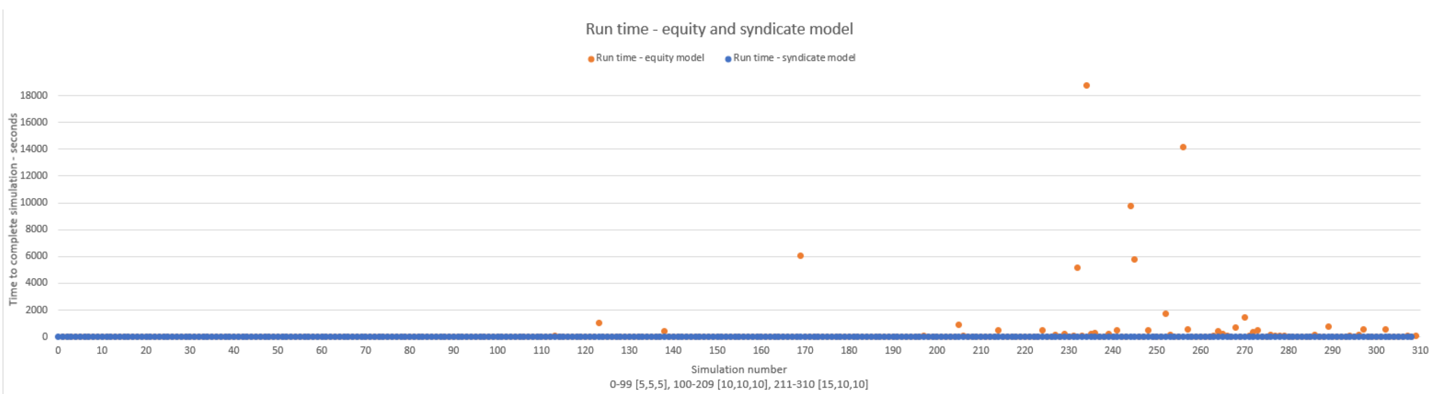

FIGURE 9. All Simulations

As one can see, there is extreme variation in the times it takes to complete the equity simulations. This staggering discrepancy is why we have limited our focus of the results to the three simulation sizes listed above, [5,5,5], [10,10,10], and [15,10,10]. While data for larger sizes exists, and are included in this paper, those data points are from smaller simulations. The complete list of average times is presented in the chart below.

Sample Size	Average run time - syndicate model (sec)	Average run time - equity model (sec)	Delta
5,5,5	0.017783685	0.155541496	0.137757812
10,10,10	0.176804026	85.18156917	85.00476514
15,10,10	0.298722172	664.2305109	663.9317888
5,10,15*	0.10651021	1.05272975	0.94621954
15,15,15*	0.84802413	629.6023943	628.7543702
10,20,30*	1.121283054	2.657177448	1.535894394
25,10,20*	0.974376917	1412.861234	1411.886857
15,30,40*	178.2820818	245.6341405	67.35205865
20,10,10*	0.377670765	44.89071059	44.51303983
50,50,50**	1482.230456		

*This data is from a single simulation, not a true average

**This data is from 70 simulations

Further research is needed to determine a relationship between changes in each of the 3 parameter sizes (bins, items, syndicates) and total computation time. The size of problem which was able to generate 100 simulations was much smaller than anticipated, and lead to the disjointed nature of the presented results.

6.3. **Inequality index measure.** A key focus of this paper is the relationship between the inequality index, and the total spend on bins for each model. Ultimately, we want to know if including equity in the objective function will drastically increase the cost. We find this is not the case. on average for the [5,5,5] model, there is a 0.60% increase in spend on bins, compared to a 5.27% increase in equity. This trend is mirrored by the [10,10,10] model with a 0.53% increase in spend, but a 3.13% increase in equity. Lastly, the [15,10,10] model has a 0.50% increase in spend, with a 1.00% equitable increase. The full results are presented in the table below. They represent the difference between the syndicate model, and the equity model.:

Sample Size	Avg ineq. index - syndicate model	Avg ineq. index - equity model	Delta	%Delta
5,5,5	8.684478454	8.249869373	0.434609081	5.27%
10,10,10	14.67870295	14.23254598	0.446156968	3.13%
15,10,10	19.65028696	19.45660276	0.193684203	1.00%
5,10,15*	11.75336921	10.11577305	1.637596159	16.19%
15,15,15*	17.37367455	17.37367358	9.66885E-07	0.00%
10,20,30*	18.07494853	17.80249123	0.272457296	1.53%
25,10,20*	33.59140914	33.59140647	2.66799E-06	0.00%
15,30,40*	20.21192275	20.00490556	0.207017187	1.03%
20,10,10*	22.14025204	22.14025172	3.17243E-07	0.00%

Sample Size	Avg. spend on bins - syndicate model	Avg. spend on bins - equity model	Delta	% Delta
5,5,5	10.07	10.13	-0.06	-0.60%
10,10,10	15.58715596	15.66972477	-0.082568807	-0.53%
15,10,10	15.93	16.01	-0.08	-0.50%
5,10,15*	14.6	14.8	-0.2	-1.37%
15,15,15*	18	18	0	0.00%
10,20,30*	27	28	-1	-3.70%
25,10,20*	15	15	0	0.00%
15,30,40*	33	33	0	0.00%
20,10,10*	12	12	0	0.00%

For further illumination, the graphs of these results are presented after the table.

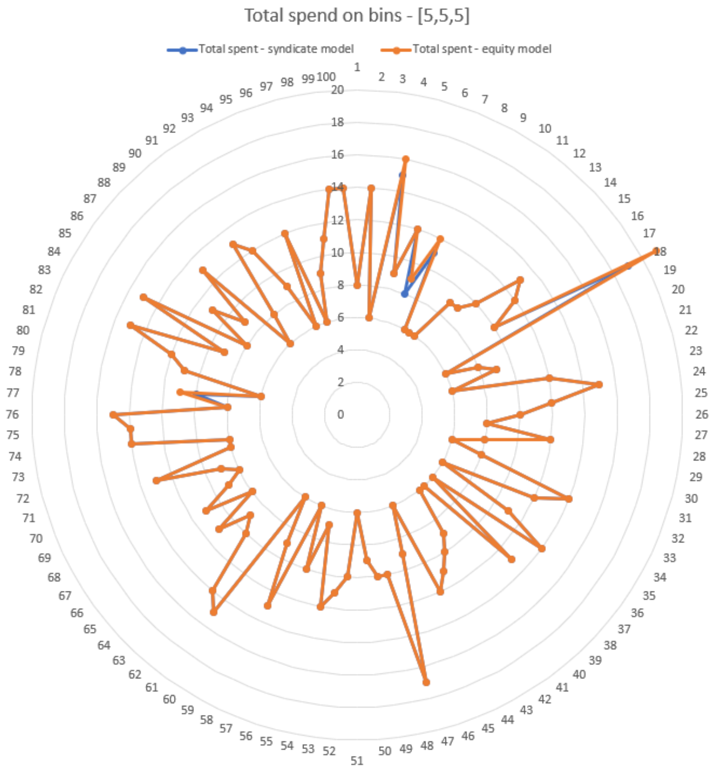

FIGURE 10. Total spent on bins [5,5,5]

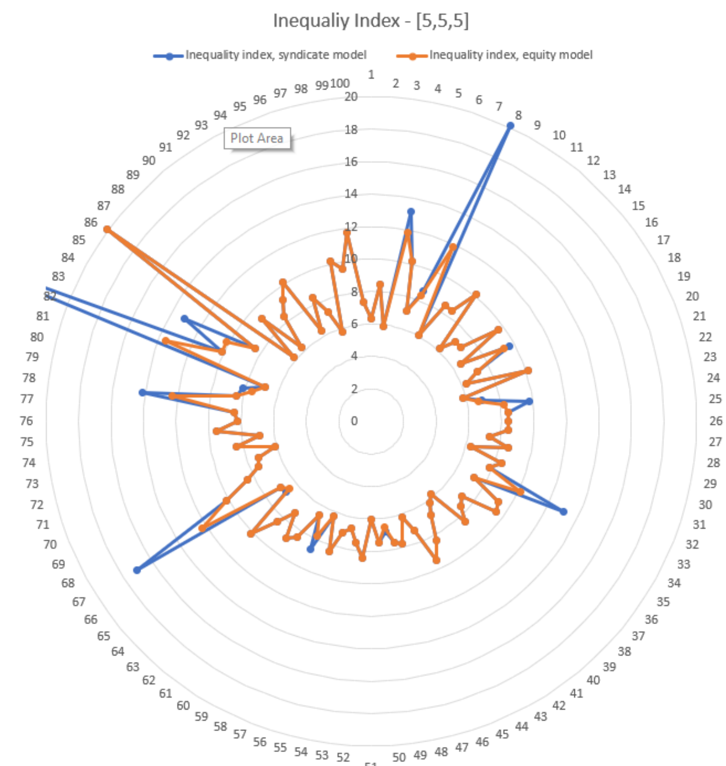

FIGURE 11. Inequality Index [5,5,5]

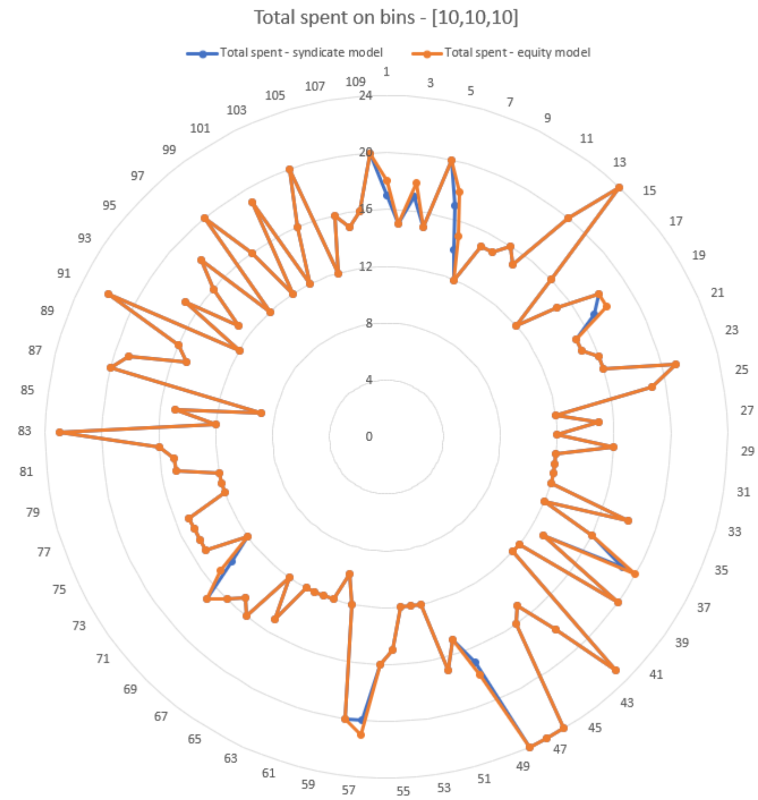

FIGURE 12. Total spent on bins [10,10,10]

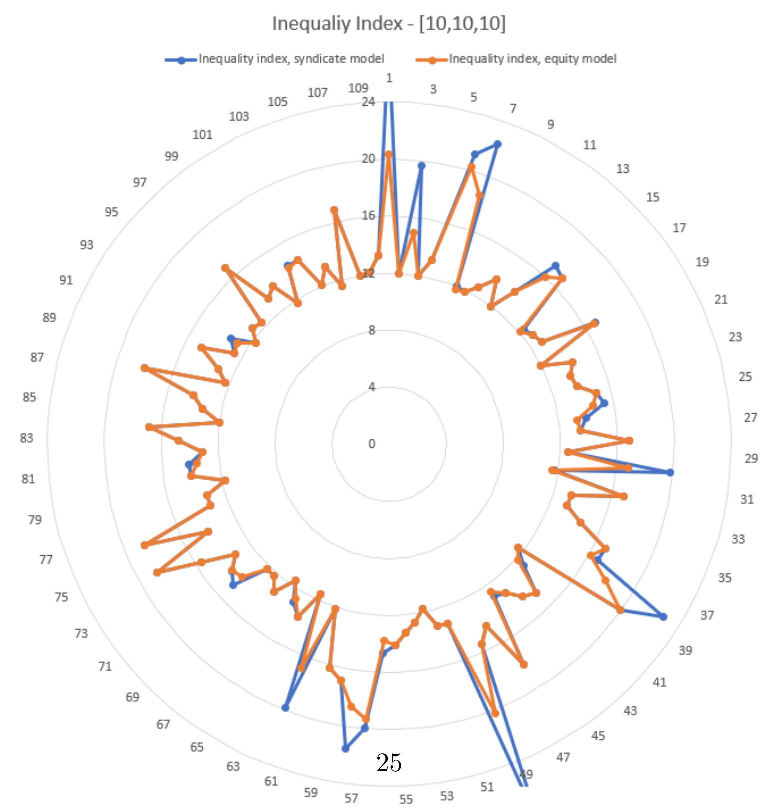

FIGURE 13. Inequality Index [10,10,10]

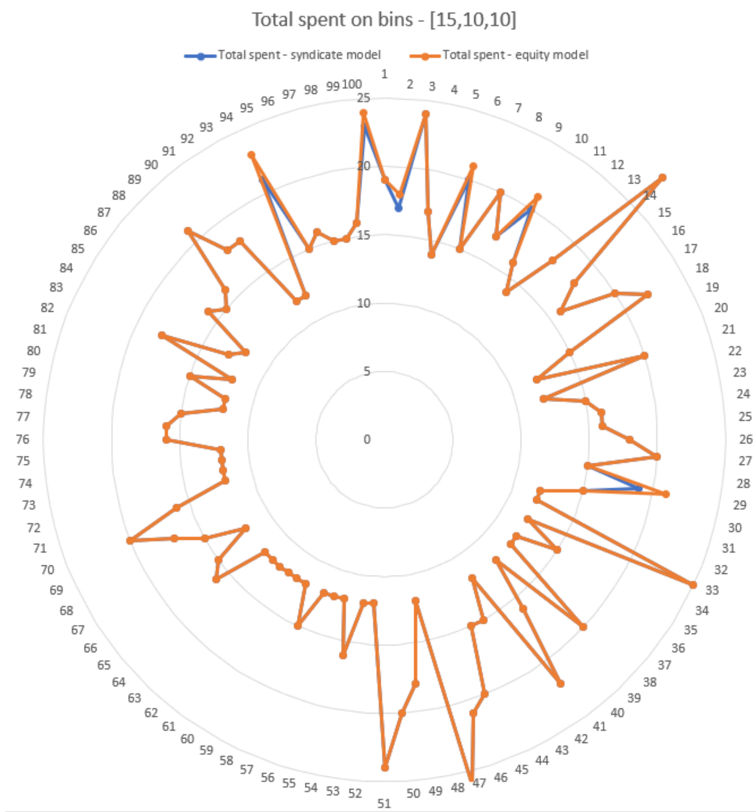

FIGURE 14. Total spent on bins [15,10,10]

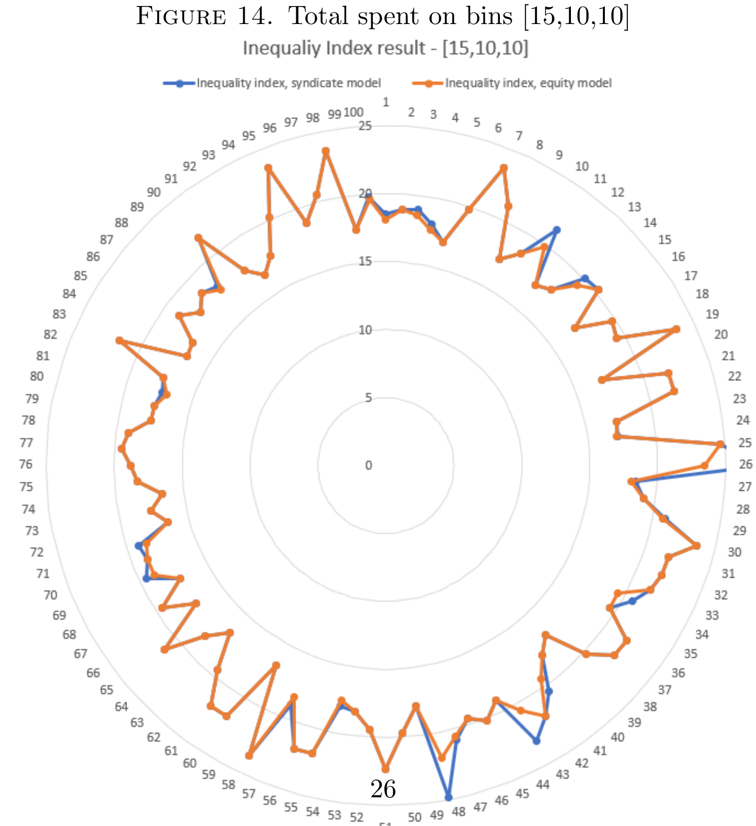

FIGURE 15. Inequality Index [15,10,10]

7. Conclusion

In this paper we have shown how to incorporate equity into another O.R. problem, the bin packing problem. The resulting non-linear model yields results that suggest, for small instances, there is a larger benefit of equity than a penalty to cost. that benefit appears to diminish as the sample size increases.

For the full code, see [Showalter-Castorena, 2008]

References

[Adger, 2000] Adger, W. N. (2000). Social and ecological resilience: are they related? *Progress in Human Geography*, 24(3):347–364.

[Apache, 2023] Apache (2023). Apache homepage.

[Atkinson, 1970] Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3):244–263.

[Blackwood and Lynch, 1994] Blackwood, D. and Lynch, R. (1994). The measurement of inequality and poverty: A policy maker's guide to the literature. *World Development*, 22(4):567–578.

[Ekici, 2022] Ekici, A. (2022). Variable-sized bin packing problem with conflicts and item fragmentation. *Computers and Industrial Engineering*, 163:107844.

[Erbayrak et al., 2021] Erbayrak, S., Özkır, V., and Mahir Yıldırım, U. (2021). Multi-objective 3d bin packing problem with load balance and product family concerns. *Computers and Industrial Engineering*, 159:107518.

[Exec. Order No. 13648, 2021] Exec. Order No. 13648, . (2021). *Executive Order on Diversity, Equity, Inclusion, and Accessibility in the Federal Workforce*.

[Foundation, 2012] Foundation, R. S. (2012). Housing and the great recession. `https://furmancenter.org/files/publications/HousingandtheGreatRecession.pdf`.

[Google, 2023] Google (2023). Google scholar profile. `https://scholar.google.com/citations?user=3vx_7UgAAAAJ`.

[Gopinath, 2020] Gopinath, G. (2020). The great lockdown: Worst economic downturn since the great depression.

[Horton, 2021] Horton, D. (2021). Health risk, inequality indexes, and environmental justice. Master's thesis, University of Colorado Denver.

[Logan et al., 2021] Logan, T., Anderson, M., Williams, T., and Conrow, L. (2021). Measuring inequalities in urban systems: An approach for evaluating the distribution of amenities and burdens. *Computers, Environment and Urban Systems*, 86:101590.

[Maguire and Sheriff, 2011] Maguire, K. and Sheriff, G. (2011). Comparing distributions of environmental outcomes for regulatory environmental justice analysis. *International journal of environmental research and public health*, 8(5):1707–1726.

[Mandel, 2023] Mandel, J. (2023). Ccm guide. `https://ccm-docs.readthedocs.io/en/latest/`.

[NSF, 2023] NSF (2023). Award abstract no. 2019089 cc compute: Accelerating science and education by campus and grid computing.

[Ruth, 2019] Ruth, T. (2019). Equity tree. `https://designintech.report/2019/03/09/design-in-tech-report-2019/`.

[Sheriff and Maguire, 2020] Sheriff, G. and Maguire, K. B. (2020). Health risk, inequality indexes, and environmental justice. *Risk analysis*, 40(12):2661–2674.

[Showalter-Castorena, 2008] Showalter-Castorena, Z. (2008). Bin packing repo. `https://github.com/Z-SC/Bin_Packing_Problem`.

[Witteman et al., 2021] Witteman, M., Deng, Q., and Santos, B. F. (2021). A bin packing approach to solve the aircraft maintenance task allocation problem. *European Journal of Operational Research*, 294(1):365–376.