BENCHMARKING BUILDING ENERGY IN THE MULTIFAMILITY INDUSTRY: A DATA ENVELOPMENT ANALYSIS (DEA) MODEL

A Dissertation
Presented to
The Academic Faculty

by

Jun Wang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
May 2017

BENCHMARKING BUILDING ENERGY IN THE MULTIFAMILY INDUSTRY: A DATA ENVELOPMENT ANALYSIS (DEA) MODEL

Approved by:

Dr. Baabak Ashuri, Advisor
School of Civil and Environmental Engineering / School of Building Construction
Georgia Institute of Technology

Dr. Eric Marks
School of Civil, Construction and Environmental Engineering
University of Alabama

Dr. Iris Tien
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Mohsen Shahandashti
School of Civil Engineering
University of Texas

Dr. Xinyi Song
School of Building Construction
Georgia Institute of Technology

Date Approved: March 23, 2017

To my parents

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Baabak Ashuri, who has always been very supportive and thoughtful, both academically and professionally. I would not be where I am today without his generous support. I would also like to thank all the other members of my doctoral committee, Dr. Iris Tien, Dr. Xinyi Song, Dr. Eric Marks, and Dr. Mohsen Shahandashti for their valuable guide and comments, which have helped me improve this dissertation a lot.

Much of my success in graduate school at Georgia Tech can be attributed to my fellow laboratory colleagues: Nipesh Pradhananga, Kyungki Kim, Aaron Costin, Jian Lu, Saman Yarmohammadi, Yang Cao, and Jianli Chen for providing me with help and suggestions in different stages of my Ph.D. study; and my friends in the QCF program: Nan Xie, Xi Wang, Hongli Jiang, and Chuanfeng Guo for spending one and half years together on various coursework and projects.

I am indebted to my parents for their continuous support and enlightenment throughout my Ph.D. study and entire life. I am also thankful for the motivation and encouragement from my girlfriend that enables me to achieve a series of accomplishments that I would not have done by myself.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

bEQ	building Energy Quotient
Btu	British thermal units
CBECS	Commercial Building Energy Consumption Survey
CDD	Colling Degree Day
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
DRS	Decreasing Returns to Scale
EIA	U.S. Energy Information Administration
EPI	Energy Performance Indicator
HDD	Heating Degree Day
IEA	International Energy Agency
IRS	Increasing Returns to Scale
OECD	Organization for Economic Co-operation and Development
RTS	Returns to Scale
TDD	Total Degree Day
VRS	Variable Returns to Scale

SUMMARY

Building energy benchmarking, offering initial building energy performance assessment, is a crucial tool for decision makers and facility managers to promoting the efficient use of energy among different properties. Traditional benchmarking models are mostly constructed in a simple benchmark table, comparing basic statistics of energy use of different properties. But they are very often subject to human judgement and are not capable of dealing with complex situations when multiple inputs and outputs are involved. Later on, linear regression model is utilized for building energy benchmarking, but it is still limited due to its various assumptions and the uncertainty of its prediction power. Recently, data envelopment analysis (DEA) has been utilized for benchmarking building energy, but existing DEA models have not been utilized to its optimum potential and are subject to limitations such as high sensitivity to outliers.

This research intends to propose an integrated approach for building energy benchmarking analysis in the multifamily industry. DEA model will be chosen in this research as it has been understudied despite its possibilities. A systematic peer-wise multifamily building energy benchmarking model based on the DEA method is the expected outcome of this research. The proposed model is expected to be capable of selecting appropriate variables to be included in the model, remediating errors in the dataset, considering weather impact on building energy consumption, and detecting outliers that may distort the final efficiency score.

This research intends to build on and contribute to the existing body of knowledge for building energy benchmarking, filling in the gaps of the knowledge in the existing DEA

building energy benchmarking method. The scope of this research is multifamily properties from different geographical regions in the United States. The proposed research has the potential to improve energy consumption by ranking properties based on different efficiency scores. Research deliverables are expected to provide decision makers and facility managers with the crucial information for building energy improvement.

CHAPTER 1. INTRODUCTION

1.1 Background

The global contribution from buildings towards energy consumption has been steadily increasing by approximately 20-40% in developed countries, exceeding other major sectors (Pérez-Lombard, 2008). The concept of energy certification for buildings was emerged in the early 1990s with the overall objective of saving energy consumption without compromising comfort, health and productivity levels (Pérez-Lombard, 2009).

It is argued that 30% or more energy usage is reduced in businesses by effective energy management practices, including assessing energy performance, setting energy saving goals, and regularly evaluating progress (Energy Star, 2008). Energy performance benchmarking of buildings is an integral part of this effort. For instance, Energy Star certification of buildings showed that comparing the energy use of buildings with other buildings nationwide help identify the opportunities of potential saving and the best practices that can be replicated (Energy Star, 2008).

Building energy benchmarking is required for adopting an energy certification scheme, promoting energy efficiency, and reducing energy consumption. It demonstrates the current level of consumption, the value of potential improvement, and the prospects for additional savings (EPBD, 2003). It promotes efficient energy consumption in the real estate market, identifies energy efficiency measures, and supports regulations of building efficiency (ASHRAE, 2015). It also helps to understand the opportunities lost by low energy performance, as well as the potential benefits of enhancing energy efficiency.

A lack of systematic building energy benchmarking method exists for the multifamily industry. Currently, most research is conducted based on examples in other sectors or industries, and very few research has bene conducted in the multifamily industry. The data of the existing benchmarking models are not enough pre-processed, often resulting in operating on garbage-in-garbage-out mechanism. That means if the input variables are carefully selected and outliers are not detected for the model, irrespective of the detail of the model, output will not be informative. The models themselves are not tailored towards the full consideration of relevant factors that may impact building energy. The generated results are not well explained, making the end users unaware of how to read and understand the results and identify potential areas of improvements. Gaps in the current body of knowledge for building energy benchmarking are expected to be filled in by this thesis.

1.2 Dissertation Organization

This research aims to provide a method for building energy benchmarking with the focus on the multifamily industry using DEA model. Table 1 provides a brief summary of the contents of each chapter.

Table 1 – Title and description of each dissertation chapter

Chapter	Descriptions
1. Introduction	This chapter introduces background of this research.
2. Literature Review	This chapter reviews the multifamily building energy consumption and existing building energy benchmarking methods, both industrially and academically.
3. Objectives, Scope, and Hypothesis	This chapter discusses the objective, scope, and corresponding hypothesis of this research.

(Table 1 continued)

4.	Research Methodology	This chapter elaborates the methodology of this research in details, including variable selection, error remediation, model formulation, outlier detection, and efficiency analysis
5.	Results and Interpretations	Based on the methodology elaborated in previous chapter, this chapter delivers the results and provides interpretations for those results
6.	Conclusions	This chapter summarizes the results findings, addresses the limitations, and discusses future research

CHAPTER 2. LITERATURE REVIEW

Each multifamily property is often characterized by a series of unique features, such as location, age, number of buildings, number of occupants, occupancy rate, etc. The variety of features of among properties can create a wide range of energy consumption levels. This research focuses on developing and testing a systematic building energy benchmarking framework to potentially improve the energy consumption by providing decision makers and facility managers with meaningful information.

The following literature review covers current energy consumption level of the multifamily industry in the United States. The review also discusses both industrial best practices and current academic research of building energy benchmarking methods, none of which is developed for the multifamily industry. Limitations of current methods are summarized and a research needs statement is derived from the review.

2.1 Multifamily Building Energy Consumption in the U.S.

Energy continues to be a world-wide issue after decades due to consistently growing consumption on yearly basis and limited amount of production. In 2010, the world primary energy consumption was 514 quadrillion British thermal units (Btu). The five largest consuming countries in that year were China, U.S., Russia, India, and Japan, and they consumed 19.4%, 18.9%, 5.9%, 4.6%, and 4.3% respectively. OECD Europe consumed 15.5%, and the all other countries consumed the rest 31.4% (EIA, 2016). Figure 1 summarizes the proportion of world primary energy consumption for major consumers.

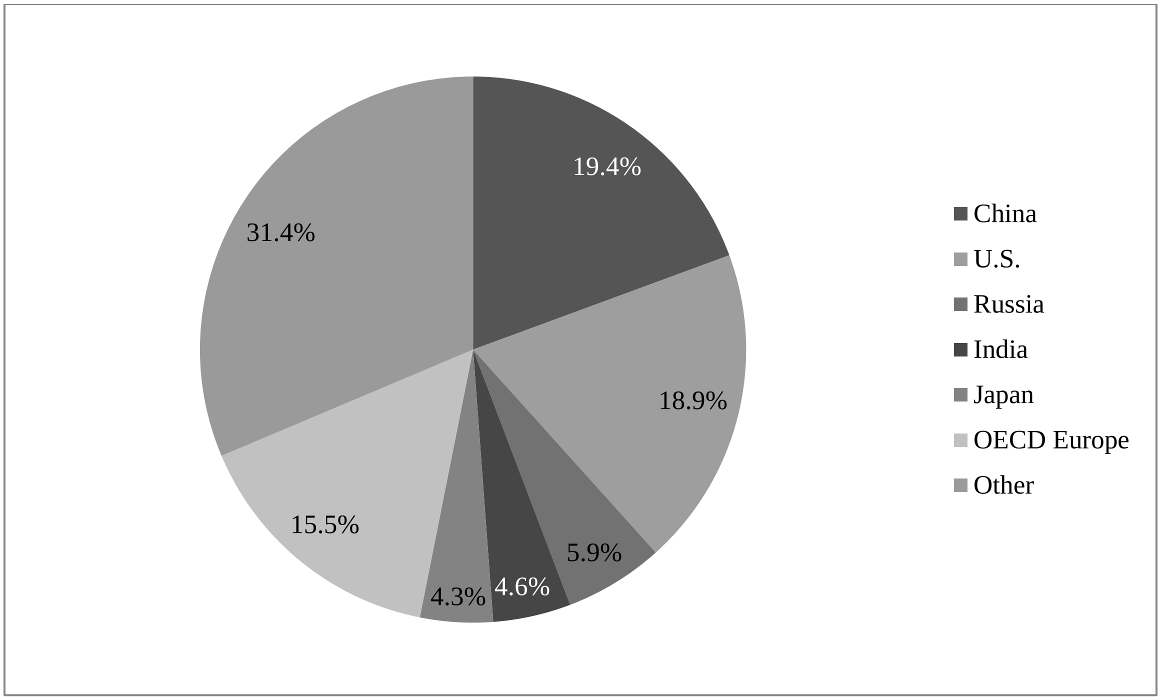

Figure 1 – Proportion of world primary energy consumption in 2010

The primary energy consumers in the U.S. can be categorized into four categories: the residential building sector, the commercial building sector, the industry sector, and the transportation sector. Among that 18.9% of world primary energy consumption, or approximately 97.5 quadrillion Btu, consumed in the U.S., the residential building sector counted for 22.5%, the commercial building sector counted for 18.6%, the industry sector counted for 30.8%, and the transportation sector counted for the rest 28.1% (PNNL, 2012). Figure 2 summarizes the proportion of U.S. primary energy consumption for different sectors.

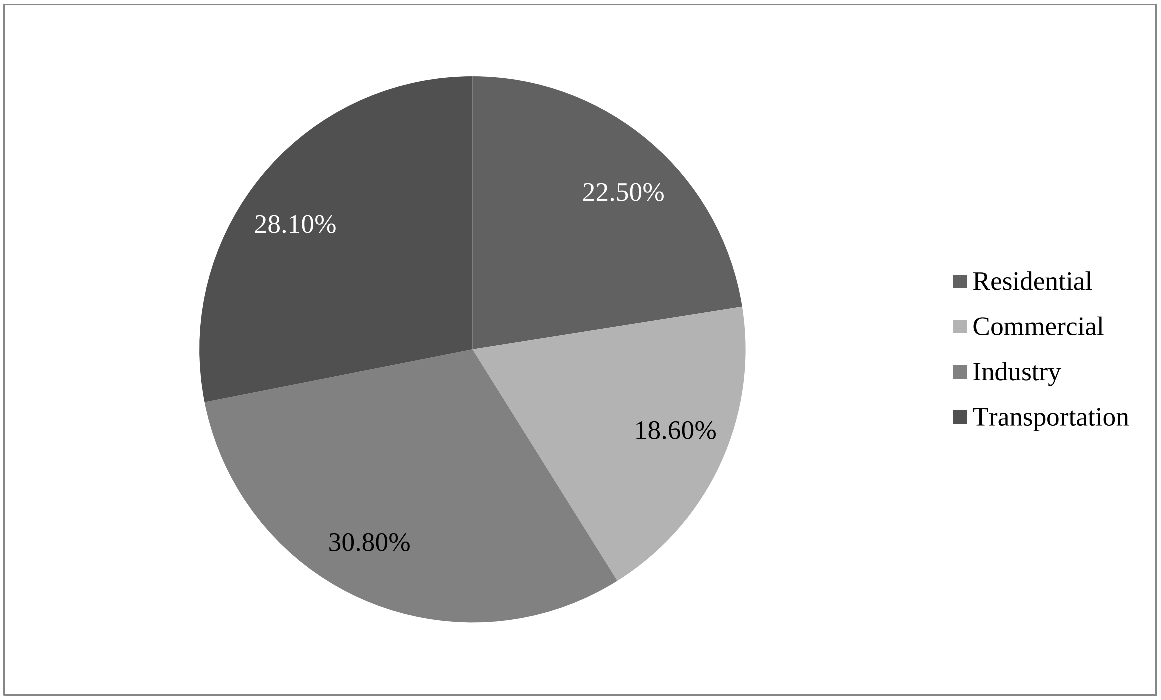

Figure 2 – Proportion of U.S. primary energy consumption in 2010

There are three different industries for primary energy consumption in the residential sector in the U.S.: the single-family industry, the multifamily industry, and mobile homes. In 2005, the residential building sector consumed 21.54 quadrillion Btu, out of which the single-family industry consumed 80.5%, the multifamily industry consumed 14.9%, and mobile homes constitutes consumed the rest 4.6%. Figure 3 summarizes the proportion of the residential sector primary energy consumption for different industries (PNNL, 2012).

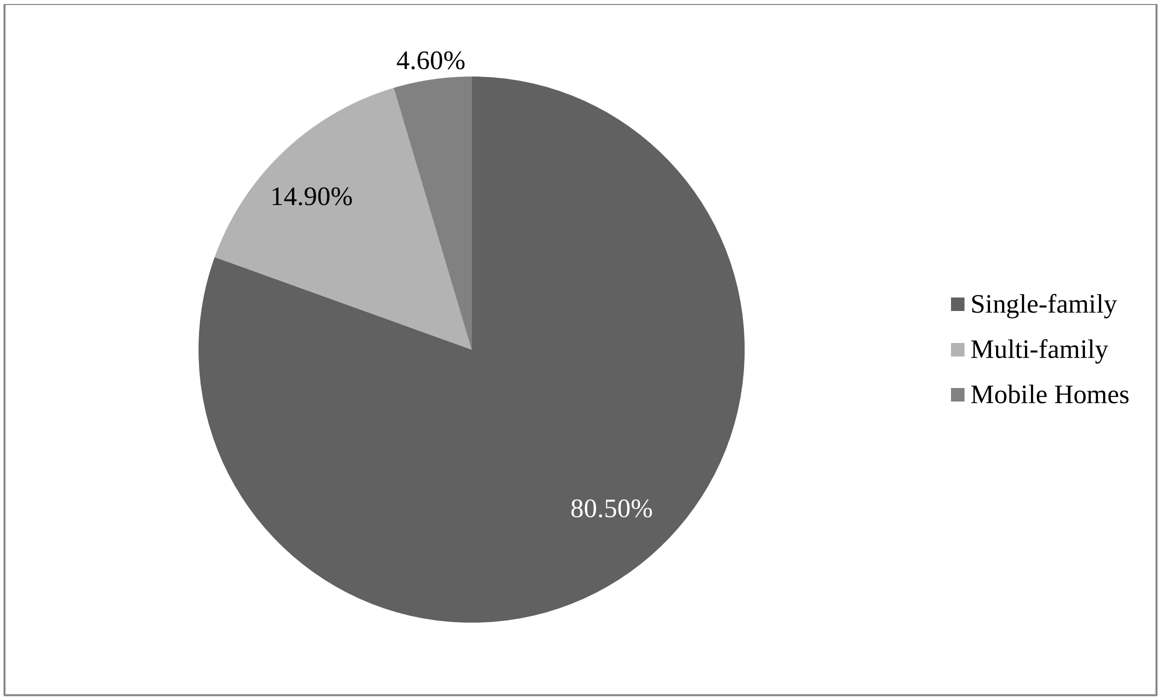

Figure 3 – Proportion of the residential sector primary energy consumption in 2005

Due to the increasing awareness of energy saving and advantage of technology innovation, there is a clear trend toward increasing energy efficiency in the residential building sector. Homes built between 2000 and 2005 utilized 44.7 thousand Btu per square foot of heated floor space, which is 14% less than homes built in 1980s and 40% less than homes built before 1950 (PNNL, 2012). Figure 4 summarizes the energy efficiency in the residential sector by vintage.

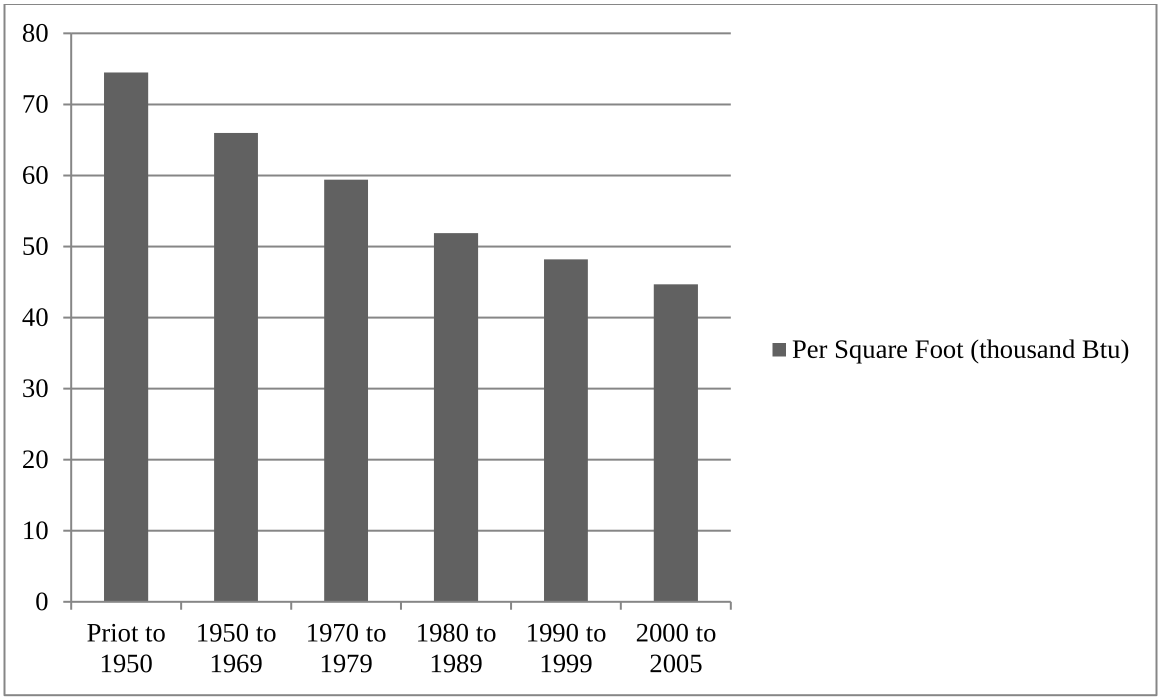

Figure 4- Energy efficiency level in the residential sector by vintage

Although multifamily industry is not consuming as much as energy consumed by the single-family industry, it is the least efficient industry in the residential sector in terms of energy consumed per square foot of heated floor space. In 2005, the multifamily industry consumed 78.3 thousand Btu per square foot of heated floor space, which is 41% more than that is consumed in the single-family industry. Figure 5 summarizes the energy efficiency of different industries in the residential sector.

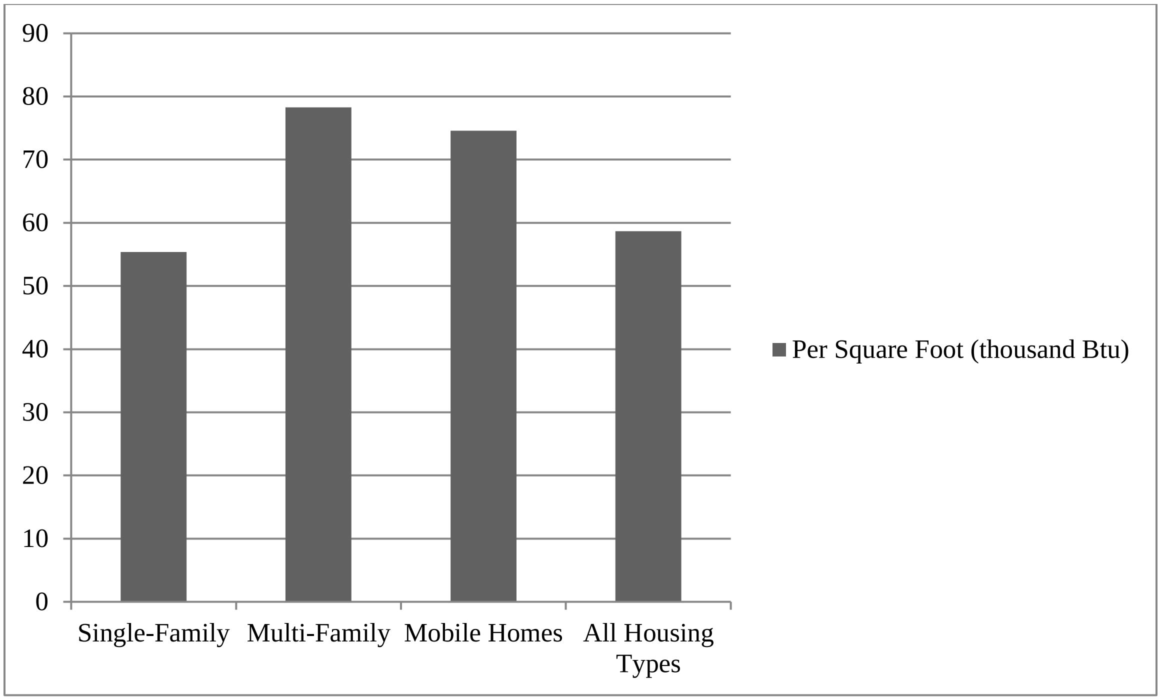

Figure 5 – Energy efficiency level of different industries in the residential sector

2.2 Building Energy Benchmarking Industrial Practices

There are two major agencies in the current market providing building energy performance benchmarking and certification issuance services: Energy Star portfolio manager and ASHRAE building energy quotient. This section reviews both methods, discusses their benchmarking methodology, and reveals the limitations of current industrial practices.

2.2.1 Energy Star Portfolio Manager

Energy Star portfolio manager provides commercial buildings with an Energy Star score on a 1 – 100 scale as an assessment of its energy performance. The score is calculated on a percentile basis: buildings receiving a score of 50 perform better than 50% of their peers; buildings receiving a score of 80 or higher are in the top quintile in terms of energy performance. The score measures the energy efficiency of a particular building compared with a peer group of buildings in the Commercial Building Energy Consumption Survey (CBECS).

To determine the score, Energy Star will compute both the actual source energy use intensity (EUI) and the predicted source EUI. Actual source EUI is the total energy consumption divided by gross floor area. Predicted source EUI is calculated by utilizing a regression equation that has been previously set up. Physical features and usage details of the building are the input of the regression equation, and the predicted EUI is the output. Efficiency ratio is later on calculated by actual source EUI/predicted source EUI. An efficiency ratio smaller than 1 means that the building is actually consuming less energy than it would have "theoretically" consumed based on the energy consumption performance of its peers. The Energy Star score is finally determined by the efficiency ratio percentile of the building.

In September 2014, Energy Star portfolio manager released Energy Star score for multifamily housing. Same methodology was adopted to calculate Energy Star score for multifamily housing, but based on different survey data from an industry survey conducted by the Federal National Mortgage Association (Energy Star, 2014c). The model utilizes independent variables such as unit density, bedrooms per unit, low rise or not, and some weather information to predict the EUI of a property, and it has 23.87% explanatory power (R^2) (Energy Star, 2014c).

2.2.2 ASHRAE Building Energy Quotient

ASHRAE building energy quotient (bEQ) is another energy rating program that provides both as-designed and in-operation energy performance assessments. Instead assignment a 1 – 100 scale score, ASHRAE provides buildings with letter-grade ratings with A+ representing zero net energy and F representing unsatisfactory. The letter-grade rating provides a particular building with a peer group energy performance comparison from the database of ASHRAE.

To determine both as-designed and in-operation ratings, ASHRAE will compute both standard EUI and metered EUI. Standard EUI is the source energy use computed using standard occupancy and operational schedules, while metered EUI is the actually measured source energy use. Both EUI's are later on compared with the median EUI of similar property type in CBECS. As-designed bEQ is calculated by standard EUI/median EUI*100, and in-operation bEQ is calculated by metered EUI/median EUI*100. The calculated bEQ can be converted to letter-grade rating using a bEQ scale definition table.

2.3 Current Building Energy Benchmarking Methods

Currently, there are three building energy benchmarking methods: simulation, statistical analysis, and data envelopment analysis (DEA). This section reviews each method, discusses their applications and limitations, and explains why a new systematic building energy benchmarking method is needed.

2.3.1 Simulation

The simulation method calculates theoretical energy consumption by setting up a mathematical model. The theoretical energy consumption is then compared with the observed energy consumption in order to evaluate the energy consumption performance (Lee, 2008). There are a number of tools to calculate the energy performance of a building through detailed dynamic simulation models, and the process generally involves developing a detailed numerical description of the building, with standard occupancy and activity templates (Hernandez et al, 2008). Federspiel et al. (2008) applied simulation method to construct a typical building energy model as the benchmark that represented the minimum amount of energy required to meet a set of basic functional requirements of laboratory buildings, and compared the actual energy consumption with the benchmark. Carriere et al. (1999) utilized the U.S. Department of Energy building simulation software (DOE-2 model) to study the design and efficient operation of HVAC systems in commercial buildings for potential energy saving.

In general, the simulation method reveals the ideal energy consumption of a building or the energy consumption with standardized weather and operating conditions (Olofsson et al., 2004). Although simulation is one of the most popular methods to study the effect of different factors on building energy use, its application for developing a benchmarking system is limited (Chung et al, 2006). The simulation method cannot be commonly used for existing buildings due to the difficulty of collecting simulation variables, such as the heat conductivity of walls and the properties of building materials. (Lee and Lee, 2009).

2.3.2 Statistical Analysis

There are three types of statistical analysis for building energy benchmarking: simple statistics method, normalization ranking method, and regression analysis method.

The simple statistics method gathers the energy consumption data of a group of properties and calculates several common statistics for comparison. Sharpe utilized the average and the median of office building energy consumption to establish an energy-efficiency benchmarking framework, based on the data collected in CBECS. It is found that medians are more reliable comparators than averages because averages can be strongly influenced by a small number of buildings with excessive energy consumption, especially when the sample is small (Sharpe, 1996). Besides, the information that can be conveyed when averages and medians are used as the benchmark is very limited – the energy efficiency of an individual building can be either above or below the benchmark (Wu et al., 2010).

The normalization ranking method incorporates the concept of EUI, often calculated by normalizing the energy usage with respect to the floor area. It identifies outperformers and underperformers by simply ranking the EUI of each building among the sample group to determine its corresponding rating of energy consumption performance. Buildings with EUIs in the best quartile are termed "Good Practices" and are set as target for other buildings to emulate (Bordass, 2005). This method has also been used to evaluate the energy consumption performance of commercial buildings (Birtles and Grigg, 1997). Later on, curve of cumulative percentile distribution of normalized EUI, also known as the benchmarking curve, is also on implemented to ranking building energy consumption performance on a more granular level (Wu et al., 2010). ASHRAE utilizes the ratio of both standard and metered EUI to median EUI to calculate as-designed bEQ and in-operation bEQ, where median EUI is estimated based on CBECS (ASHRAE 2009). Despite the simplicity of implementation, the normalization ranking method is limited in scope. It cannot normalize other factors related to the building energy efficiency such as property age, number of occupants, etc., which may cause energy usage in specific buildings to be higher (or lower).

Regression analysis method utilizes the EUI of each building as dependent (response) variable and several other building characteristics, such as building age and internal floor area as independent (explanatory) variables. The objective is to construct a multivariate regression model of the independent variables to explain variations in EUIs as the response variable. The developed regression model will be used to predict EUI given certain values of explanatory variables of each building. Eventually, the predicted EUI will be compared with the actual EUI of each building to construct a benchmarking table. For instance, Energy Star Portfolio Manager defines the percentile distribution of the ratio of actual source EUI to predicted source EUI as the efficiency ratio, and utilizes that for benchmarking (Energy Star, 2014a). Chung et al. (2006) developed a multiple linear regression model for supermarket buildings in Hong Kong to predict normalized EUI with standardized values of explanatory variables as the input, deriving a benchmarking table for end-users. Wu et al. (2010) also utilized a multiple linear regression model to benchmark energy efficiency of hotel buildings with different operation standards in Singapore.

Although the regression analysis method is commonly used for building energy benchmarking, it is subject to several significant limitations:

- The assumption that regression errors are normally distributed may not hold given the large variety of building characteristics. If the normal distribution assumption is violated, then the percentile-based benchmarking table will become unbalanced. The main problem with unbalanced benchmarking is that the analysis results will become unreliable to use as building energy efficiency rankings become very sensitive to small changes in the EUI.

- The assumption that explanatory variables have no or little multicollinearity may not hold given the high correlation between building characteristics. For example, the number of apartments in a building can be highly correlated with the number of washing machines in a building.
- The predictive power of the linear regression model may be uncertain. The relationship between the predicted EUI and building characteristics (model coefficients) is not linear and it can change over time due to the variation of the sample group.
- The wellness of the fitted regression model may be neglected. The ratio of actual EUI to predicted EUI measures both the actual energy efficiency and variation of outcomes that are not explained by the model. For example, the regression baseline model estimated by Energy Star for multifamily housing in the United States has 23.87% explanatory power (i.e., adjusted R^2 = 22.66%) indicating that there is a large modeling error due to unexplained factors or data errors (Energy Star, 2014 c).

2.3.3 Data Envelopment Analysis

DEA is a data-oriented approach for evaluating the performance of a set of homogeneous entities called decision-making units (DMUs) (Cooper et al., 2011). DEA is a peer-to-peer comparison method that evaluates the relative performance of a DMU in a pool of comparable DMUs. Several inputs and multiple outputs are considered in relative performance assessment. DEA utilizes linear programming technique to compute a non-parametric frontier as the benchmark to assess the performance efficiency of DMUs. In the context of building energy benchmarking, DEA treats each building as a DMU in a multiinput/multi-output environment and computes an overall optimal frontier as the benchmark from the data in hand. Buildings located on the frontier are efficient DMUs that have generated the maximum outputs for their levels of inputs, and other buildings are evaluated based on their overall performance relative to that of the buildings on the frontier. Unlike regression method, DEA does not estimate parameters for the model, but it identifies a non-parametric frontier that constitutes the outperformers of the group. The performance of one DMU in the DEA model is dependent on its relative performance compared with the frontier, i.e., the performance of other DMUs in the model.

DEA has been utilized for benchmarking building energy consumption. Önüt and Soner (2006) applied DEA method to benchmark energy usage of 32 five-star hotels based on utility billing data and identified the most energy-efficient (called “best practices”) hotels as the ones that are on the frontier. Lee (2008) collectively utilized multiple linear regression to find out the predicted EUI of units evaluated and DEA method to calculate overall energy efficiency, using the predicted EUI as output and the observed EUI as input. Lee and Lee (2009) developed a DEA model to benchmark energy efficiency of 47 government office buildings and divided the overall energy efficiency into scale factor and management factor. Grösche (2009) used data from the U.S. residential energy consumption survey (RECS) to develop a DEA model to measure energy efficiency improvements of single-family residential buildings. It was concluded that a substantial part of the variation in energy scores is due to climatic influences but households have nevertheless improved their energy efficiency. Hui and Wan (2013) employed DEA method to study the energy benchmarking of hotels in Hong Kong and showed that DEA provides a helpful benchmarking framework for understanding efficiency within an organization that uses a variety of resources to provide a complex set of services in multiple locations. Lu et al. (2014) utilized DEA method to benchmark the energy consumption of 90 multifamily properties. They calculated the energy efficiency in a time series manner for twelve months. Most recently, Wang et al. (2015) utilized a two-stage DEA method to benchmark the energy consumption of 189 one-story single-family buildings in Woodbine, Iowa, combining DEA method with Tobit regression for further efficiency analysis.

2.4 Problems and Needs Statement

By comparing all three methods mentioned above, I found that both simulation method and statistical analysis have several intrinsic limitations for building energy benchmarking, while DEA is the one with a lot of potential but is has not been fully explored. In summary, current applications of DEA method for building energy benchmarking are subject to six main problems that highly limit their applications for energy benchmarking in the multifamily sector:

1. Very little of the current research is conducted based on the context of the multifamily industry;
2. None of the existing research shows how to handle missing or incorrect variable values in the dataset. It is quite common to have missing or incorrect variables in the dataset collected by property managers. Simple removal of a record with a missing or incorrect value may result in insufficient number of data records and eminent risk of changing the shape of the efficient frontier;
3. Most current research (Önüt and Soner, 2006; Lee, 2008; Lee and Lee, 2009, Grösche, 2009; Hui and Wan, 2013) does not take into account the issue of outliers. The results of DEA are sensitive to outliers and can be misleading if

outliers exist. (Tran et al., 2010; Lu et al., 2014; Khezrimotlagh, 2015; Wang et al., 2015);

4. Several of the existing DEA models (Önüt and Soner, 2006; Lee, 2008; Lu et al., 2014; Wang et al., 2015) consider EUI as an input variable. However, one of the main assumptions in the definition of efficiency measure under the DEA formulation, the convexity axiom, may be violated if EUI is considered as an input variable (Emrouznejad and Amin, 2009);
5. None of the existing research distinguishes between controllable variables (such as number of tenants in a building) and non-controllable variables (such as weather conditions). However, it is crucial to differentiate controllable variable from non-controllable variables because: (a) Property managers simply do not have any control over the weather conditions; and (b) The weathers conditions cannot be scaled up or scaled down;
6. None of the existing research of building energy benchmarking conducts the sensitivity analysis of the efficiency scores derived from DEA model, but those scores are subject to change, and sometimes may even be very volatile.

Table 2 summarizes problems addressed by current DEA application for building energy benchmarking and problems to be addressed by the proposed research.

Table 2 – Problems of current DEA application for building energy benchmarking

Problems of current research	Önüt and Soner (2006)	Lee (2008)	Lee and Lee (2009)	Grösche (2009)	Hui and Wan (2013)	Lu et al. (2014)	Wang et al. (2015)	**Proposed research**
1. Type of building								
1-1. Hotel	●				●			
1-2. Government office		●	●					
1-3. Residential – single-family				●			●	
1-4. Residential – multifamily						●		●
2. Remediate missing/incorrect values								●
3. Detect and remove outliers						○	○	●
4. Misuse EUI as input variable	●	●				●	●	
5. Consider non-controllable variables								●
6. Conduct sensitivity analysis								●

(●: fully addressed; ○: partially addressed)

The need for a DEA model that benchmarks building energy in the multifamily industry, therefore, still exists. The model needs to be able to handle missing/ incorrect values in the dataset, detect and remove outliers from influencing end results, exclude the usage of EUI, and consider controllable and non-controllable variables differently. Detailed efficiency analysis needs be conducted and corresponding interpretations needs to be given. Additionally, sensitivity analysis is also needed to measure the stability of results given by the DEA model.

CHAPTER 3. OBJECTIVES, SCOPE, AND HYPOTHESIS

In order to understand the goal, purpose and methodology of this research, the objective, scope and hypothesis must be defined. In the subsections to follow, each of these research components are discussed.

3.1 Objectives

The major objective of this research is to create a new DEA-based approach for benchmarking energy efficiency in buildings in the multifamily sector to address the major limitations of existing DEA models. To achieve this objective, several necessary secondary objectives are listed:

- To find a method that remediates missing or incorrect values for instances in the dataset
- To establish a mechanism that accurately and effectively detects outliers in the dataset
- To select appropriate variables to be included in the DEA model and provide justification for the selection
- To build up a DEA model that differently handles controllable variables and non-controllable variables
- To quantitatively measure the stability of efficiency scores of each DMU across the entire period

3.2 Scope

The proposed research will only focus on the multifamily industry of the residential sector, other than other industries of the residential sector such as single-family or mobile homes, or other sectors such as commercial or industrial sector. The data of both energy consumption and building characteristics will be provided by a third-party organization, so the proposed research will not focus on data collection, but on data preparation, model formulation, and results interpretation. The only data needs to be collected in this research is the weather information of the place where each property is located, which can be accessible from publicly available database The energy consumption analyzed in this research is the total energy consumption for each property. The research will only focus on DEA method, other than several other methods as summarized in the literature review part.

3.3 Hypothesis

After reviewing existing research and current practices regarding building energy benchmarking, the following hypotheses were generated by the researcher and will be testing using the research methodology described in the next chapter:

- A method can be found to remediate missing or incorrect values for instances in the dataset
- A mechanism can be established to accurately and effectively detects outliers in the dataset
- Appropriate variables for the DEA model can be selected and justifications can be provided accordingly
- A DEA model that handles controllable variables and non-controllable variables differently can be built up

- The stability of efficiency scores of each DMU across the entire period can be quantitatively measured

CHAPTER 4. RESEARCH METHODOLOGY

This chapter defines the foundational components for this research and describes in detail the methodology for each of the research components. Section 4.1 reviews the variable selection for DEA model in past research, and justifies how to choose appropriate without violating the convexity assumption of DEA model. Section 4.2 discusses the method of remediating data errors in this research. Section 4.3 explains how to take weather influence into consideration for building energy benchmarking. Section 4.4 covers the model formulation using both constant return to scale (CRS) model and variable return to scale (VRS) model, and discusses how to differentiate controllable variables and non-controllable variables in DEA model. Details of detecting and removing outliers in dataset will be elaborated in Section 4.5. Section 4.6 introduces and explains the three different efficiency scores that DEA model generates. Lastly, the sensitivity and stability analysis of the results of DEA model is presented in Section 4.7. Figure 6 shows the methodology flowchart of the proposed research. Each component of the flowchart is discussed in the following sections.

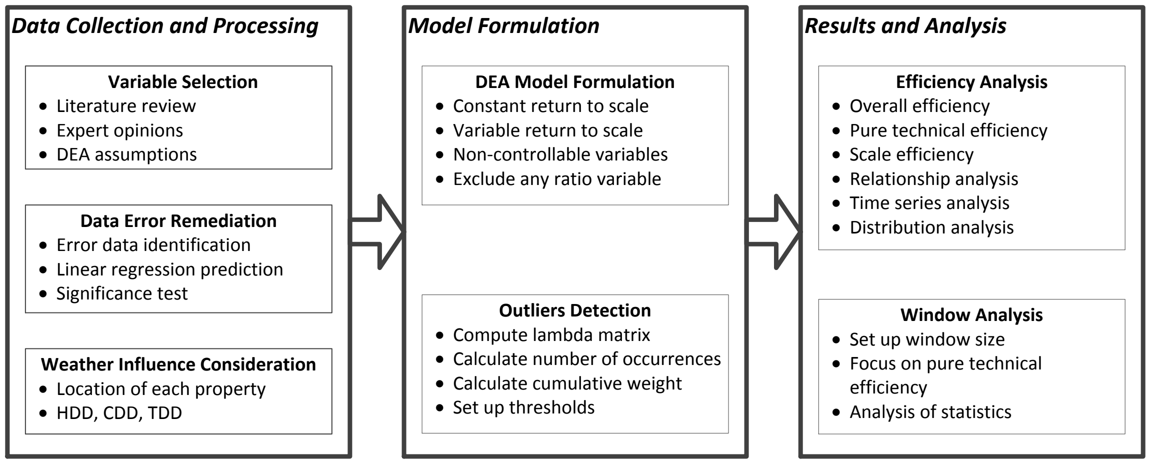

Figure 6 – Methodology flowchart of the proposed research

4.1 Variable Selection

A drawback of DEA model is that the inclusion/exclusion of variables can affect the results (efficiency scores), and there is no way to test the appropriateness of each input and output variable (Hui and Wan, 2013). It makes selecting appropriate variables to be included in the DEA model extremely important, as inappropriate variable selection may lead to unreliable benchmarking results. The selection of variables is always dependent on the availability of data. Three principals should be considered in selecting appropriate variables for constructing DEA model:

1. Conduct literature review to examine the experience of other researchers on the same or similar industry;
2. Seek subject matter experts' opinions;
3. Make sure variables included would not violate any fundamental assumptions of DEA model.

These principals are described in more details below.

4.1.1 Summary of Variable Selection in Existing Research

A comprehensive literature review is conducted on variable selection for all DEA models applied to different building types when benchmarking energy efficiency. Details are shown in Table 3.

Table 3 – Summary of variables used for DEA building energy benchmarking

Variables	Önüt and Soner (2006)	Lee (2008)	Lee and Lee (2009)	Grösche (2009)	Hui and Wan (2013)	Lu et al. (2014)	Wang et al. (2015)
Input Variables							
Number of employees	●						
Energy consumption			●	●	●		
EUI	●	●					
Outdoor temperature					●		
Relative humidity					●		
Weather normalized EUI						●	●
Output Variables							
<u>Hotels</u>							
Annual total revenue	●						
Food & beverage covers					●		
Room nights					●		
Room guests					●		
<u>Building characteristics</u>							
Age of properties						●	●
Basement type							●
Buildings conditions							●
Floor area			●	●			●

(Table 3 continued)

Number of apartments						●	
Number of bathrooms							●
Number of bedrooms						●	
Number of buildings						●	
Number of fridges				●			
Number of parking lots						●	
Number of washing machines						●	
Type of AC systems							●
<u>Tenants</u>							
Number of occupants			●	●		●	●
Occupant intensity		●					
Occupancy rate	●						
Total number of guests	●						
<u>Weather Conditions</u>							
Average outdoor temperature		●	●				
Average hours of rain		●	●				
CDD				●			
HDD				●			

Three characteristics about variable selection for building energy benchmarking using DEA method can be seen from Table 1: (1) The number of input and output variables used in the existing DEA models is within a range of 4-8 with the average value of 6; (2) There is a big inconsistency in the utilization of output variables that most output variables are selected only once in the existing DEA models. The only two variables that have been utilized more than twice in the existing research are: floor area (3 times) and number of occupants (4 times); and (3) There is also big discrepancy in the utilization of input variables. Energy consumption, EUI, and weather normalized EUI have all been used in different DEA models.

The variable selection process should also incorporate the factors that are important in the perspective of property managers, and can be adjusted accordingly based on the opinions from different subject matter experts.

The output variable discrepancy can be explained by the fact that buildings in different industries may have different energy consumption features. Besides, the availability of data may also be a concern. The biggest problem here, however, is the discrepancy of the selection of input variables. Out of the seven existing DEA models cited in Table 3, three models used energy consumption (kWh for electricity) as the input, two models used floor area normalized energy consumption (i.e., EUI) as the input, and two models used weather and floor area normalized energy consumption (the weather normalized EUI) as the input.

4.1.2 The Inclusion of Ratio Variables in DEA Model

It is important to take into account that different building sizes and outdoor weather conditions have big influence on building energy consumption. However, simple normalization of energy consumption by floor area or weather conditions may violate one of the main assumptions in the definition of efficiency measure underlying DEA method: the convexity axiom (Emrouznejad and Amin, 2009).

Let's consider the following made-up example, shown in Table 4 and Figure 7, as the case when convexity axiom would be violated if EUI is included as an input variable.

Table 4 – Example of violating convexity axiom using EUI as input variable

DMU building	Energy consumption (kWh)	Floor area (ft^2)	EUI (kWh/ft^2)	Occupants (number)
B_1	300	150	2	8
B_2	135	100	1.35	5
B_3	150	200	0.75	3

As shown in Figure 8, B_1 and B_3 determines the efficient frontier, and B_2 is an inefficient DMU in this case. According to convexity axiom of DEA, the convex combination of B_2 and B_3, $B_{23} = \alpha B_2 + (1-\alpha)B_3, \alpha \in [0,1]$,should also be a feasible solution and stands on the right hand side (R.H.S.) of the frontier. Assume $\alpha = 0.5$, we can calculate the convex combination of B_2 and B_3 based on EUI and occupants as

$$B_{23}(EUI) = 0.5 \times B_2(EUI) + 0.5 \times B_3(EUI) = 1.05$$

$$B_{23}(Occupants) = 0.5 \times B_2(Occupants) + 0.5 \times B_3(Occupants) = 4$$

However, the actual convex combination of two buildings should be calculated as

$$B_{23}^*(EUI) = \frac{Energy\ consumption\ for\ B_{23}}{Floor\ area\ for\ B_{23}}$$

$$= \frac{\frac{1}{2}(Energy\ consumption\ for\ B_2 + Energy\ consumption\ for\ B_3)}{\frac{1}{2}(Floor\ area\ for\ B_2 + Floor\ area\ for\ B_3)}$$

$$= \frac{\frac{1}{2}(135+150)}{\frac{1}{2}(100+200)} = 0.95$$

By plotting B_{23} and B_{23}^* in Figure 8, we see that B_{23} is within the feasible solution area, but B_{23}^* is not. It is therefore concluded that the ratio variables (such as EUI) cannot be utilized directly in the DEA model due to the likelihood of violating the assumption of the model.

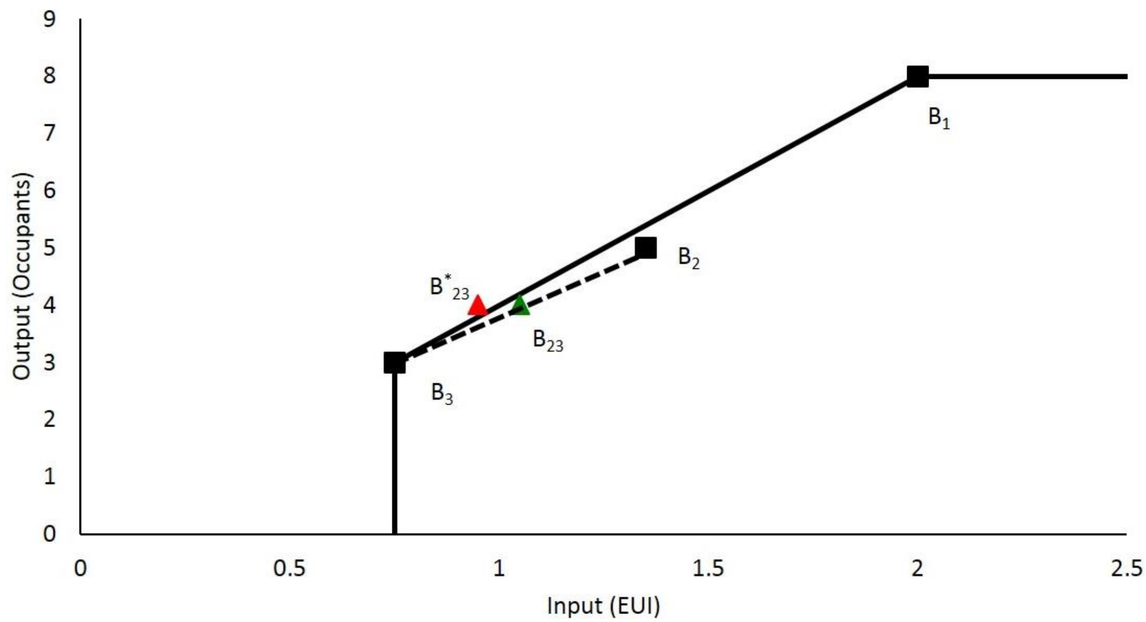

Figure 7 – Illustration of the convexity assumption of DEA model

Emrouznejad and Amin (2009) provided two solutions if any variable in the DEA model is in the form of ratio: (i) Treat the numerator and denominator separately as input or output variables in the model; or (ii) Calculate the correct convex combination of the ratio variable to be included in the model. This research chooses to model based on the first recommended solution for the following two reasons: (a) Both the numerator (energy consumption) and denominator (floor area) of the ratio variable (EUI) in this case are known and it is easy to separate them; and (b) There is no reason to treat floor area differently from several other output variables, such as number of occupants or number of washing machines. For example, if energy consumption per floor area (EUI) can be utilized as input variable, then why can energy consumption per capital (energy consumption / number of occupants) not be utilized as input variable?

4.2 Data Error Remediation

There are generally two types of errors in the data for DEA analysis: missing values or incorrect values. One of the fundamental assumptions of the original DEA method is that all the data required are available, because missing values cannot be handled by the original DEA models (Smirlis et al., 2006). Incorrect values are hazardous for DEA analysis when the values of input variables are unreasonably small or the values of output variables are unreasonably large, and either case could change the DEA frontier dramatically. For example, an apartment with more than 400 residents having only three bedrooms in the building is apparently incorrect data. However, no method of remediating data errors, such as missing or incorrect values in the dataset was proposed in the past by other research for building energy benchmarking. A simple method that has often been used is to deleting any instance with a missing or incorrect value and use the rest of data for DEA modeling, but this method is subject to two main limitations: (a) Deleting too many instances that affect the reliability of the DEA model; or (b) Dramatically changing the shape of the DEA efficient frontier by deleting potential efficient DMUs and consequently affecting the efficiency scores of the remaining DMUs.

DEA is able to locate inefficient units more powerfully when the sample size (number of instances in the dataset) exceeds the total number of output and input variables (Sherman and Gold, 1985). Three criteria need to be considered in constructing a proper DEA model: (1) The sample size should be greater than twice the product of the number of input and output variables; (2) The sample size should be greater than three times of the sum of the number of input and output variables; and (3) The total number of perfectly efficient DMUs (with score of 100) in the final results should not exceed one third of the sample size (Avkiran, 2006).

Data errors can occur on both the input variable side (i.e. energy consumption) and the output variable side (i.e. building characteristics). Multiple linear regression technique can be utilized to remediate data errors on the output variable side, as the properties features are likely to be linearly related with each other. The linear regression technique is not capable of remediating data errors on the input variable side because of the limitations of regression method discussed in Section 2.3.2. Detailed procedures are proposed as follows:

1. Delete DMUs with input variable data error or with more than one output variable data error, and keep DMUs with no error (good DMUs) or with only one output variable error;
2. Separate good DMUs from DMUs with only one output variable error;
3. Iteratively build multiple linear regression model with one output variable as dependent variable each time and all other output variables as independent variables using good DMUs
4. For each DMU with only one output variable data error, recalculate the value of that output variable using corresponding regression model built in step 3

Several concerns need to be taken into consideration when utilizing multiple linear regression to remediate data errors. First, buildings in the dataset need to be homogeneous; otherwise, the prediction power of linear regression models are limited. Second, the method remediates DMUs with only one output variable error, and is not intended to remediate DMUs with more than one output variable error or with input variable errors, both of which will not be included for further analysis. Finally, linear regression models are subject to independent variable significance test and determination of coefficient (R^2) before they can be utilized for calculation; and insignificant independent variables should not be included in the model.

4.3 Weather Influence Consideration

Because of the significant impact of weather conditions on building energy, it is critical to take into account the weather effect when benchmarking building energy consumption, particularly if buildings are from different geographical locations. Previous research has utilized total degree days (TDD) as proxy of temperature related energy consumption when benchmarking building energy consumption (Grösche, 2009; Lu et al., 2014; Wang et al., 2015).

Degree-days are a common energy accounting practice, and each degree deviation from a predefined balance point temperature is counted as a degree-day (Amato, A.D. et al, 2005). It is based on a V-shape temperature energy consumption relationship as shown in Figure 8, and energy demand is at minimum when the temperature is at balance point as the outside climatic conditions produce the desired indoor temperature (Jager, 1983; Amato et al., 2005).

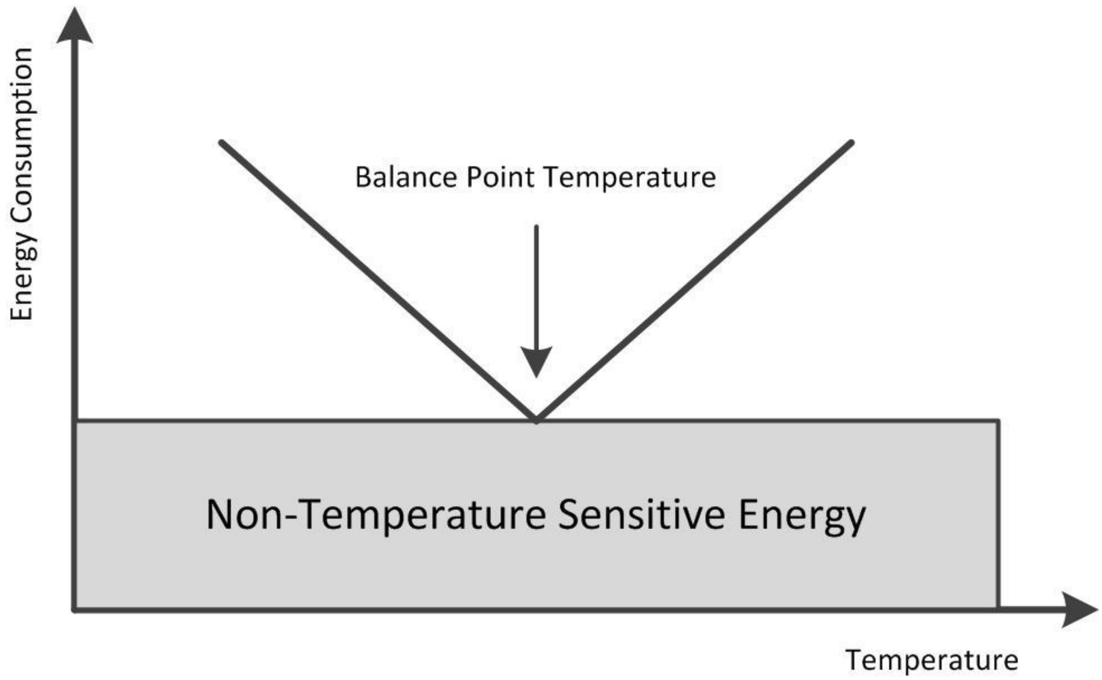

Figure 8 – Illustration of balance point temperature for degree days (Jager, 1983)

When the daily average outside temperature is below the balance point temperature, it generates heating degree days (HDD) as it requires additional energy to heat the building up. On contrast, it generates cooling degree days (CDD) when the daily average outside temperature is above the balance point temperature as it requires additional energy to cooling the building down. For example, if the balance point temperature is 65^{o}F, and the average daily outdoor temperatures of the week are 45^{o}F, 50^{o}F, 55^{o}F, 60^{o}F, 65^{o}F, 70^{o}F, and 75^{o}F, then the weekly HDD is 50 (20+15+10+5=50), and the weekly CDD is 15 (5+10=15). The TDD is defined as the summation of HDD and CDD. Most energy analyses commonly use a base temperature of 65^{o}F as the balance point threshold (Amato et al., 2005), so this research will use 65^{o}F as the balance point temperature.

4.4 DEA Model Formulation

Mathematically, DEA energy benchmarking model can be formulated following the linear programming technique suggested by Charnes et al. (1978) as

$$\begin{aligned}
&\max_{u_r, v_i} \eta_0 = \frac{\sum_{r=1}^{N} u_r y_{r0}}{\sum_{i=1}^{M} v_i x_{i0}} \\
&s.t. \\
&\frac{\sum_{r=1}^{N} u_r y_{rj}}{\sum_{i=1}^{M} v_i x_{ij}} \leq 1; \quad j = 1, \ldots, n, \\
&u_r, v_i \geq 0; \qquad r = 1, \ldots, s; \; i = 1, \ldots, m
\end{aligned} \tag{1}$$

where η_0 is the DEA efficiency score of Building 0 under consideration; y_{rj}, x_{ij} (all positive) are the known outputs and inputs of the jth building; $u_r, v_i \geq 0$ are the variable weights to be determined by the solution of this problem and are constrained to be nonnegative in order to avoid any input or output being assigned a negative weight; n is the number of buildings in the dataset; s is the total number of outputs; and m is the total number of inputs. If the solution of Model (1) is $\eta_0 = 1$, then Building 0 is considered to be 100% efficient.

Model (1) above, also known as the CCR model, is a nonlinear programming formulation of an ordinary fractional programming problem (Charnes et al., 1978), but it can be equivalently transformed into a linear programming problem as follows (Cooper et al., 2011)

$$\begin{aligned}
&\max_{u_r, v_i} z = \sum_{r=1}^{N} u_r y_{r0} \\
&s.t. \\
&\sum_{r=1}^{N} u_r y_{rj} - \sum_{i=1}^{M} v_i x_{ij} \leq 0; \quad j = 1, \ldots, n, \\
&\sum_{i=1}^{M} v_i x_{i0} = 1; \\
&u_r, v_i \geq 0; \qquad\qquad r = 1, \ldots, s; \; i = 1, \ldots, m
\end{aligned} \tag{2}$$

4.4.1 Constant Return to Scale

Since the advent of the CCR model, the economic connect of returns to scale (RTS) has been widely studied within different frameworks provided by these models (Banker et al., 2004). There are generally three forms of RTS: increasing returns to scale (IRS), constant returns to scale (CRS), and decreasing returns to scale (DRS). IRS represents outputs increase more than the proportional increase of inputs; CRS represents outputs increase proportionally as inputs increase; and DRS represents outputs increase less than the proportional increase of inputs.

To relax the computation intensity, the linear programming problem in model (2) can be transformed into its corresponding dual form as follows

$$\begin{aligned}
&\min_{\lambda_j} \theta_0 \\
&s.t. \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_0 x_{i0}; \quad i = 1,\ldots,m, \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}; \quad r = 1,\ldots,s, \\
&\lambda_j \ge 0; \quad j = 1,\ldots,n
\end{aligned} \tag{3}$$

where θ_0 is the DEA efficiency score of Building 0 that is under consideration; λ_j is the decision variable of the dual problem; y_{rj}, x_{ij}, m, s, and n would have the exactly same meaning and constraints as defined in Model (1). If the solution of Model (3) is $\theta_0 = 1$, then Building 0 is considered to be 100% efficient.

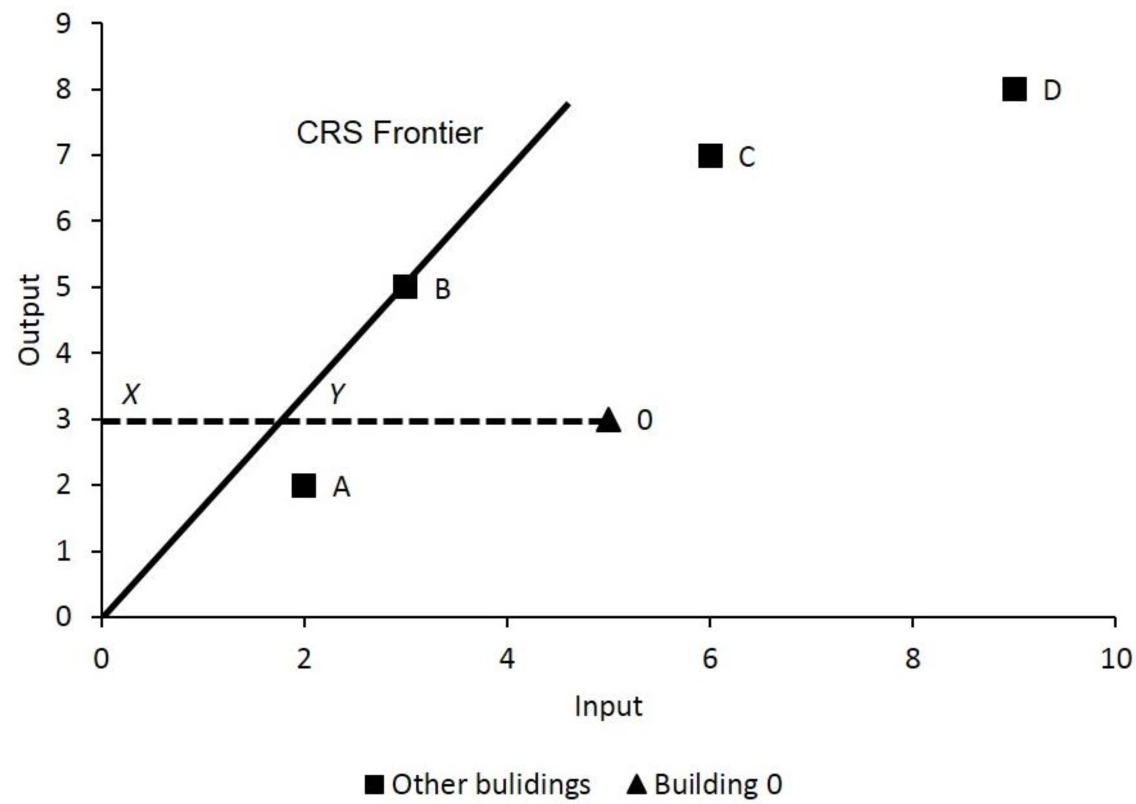

Figure 9 – Illustration of CRS DEA model

Model (3) is essentially looking for a virtual efficient building on the frontier for building 0, and compares the performance of that virtual building with building 0. Figure 9 graphically illustrates how Model (3) works in a simplified one-input-one-output scenario. Assume we have building 0 with (input, output) being $(5,3)$, and we have four additional buildings in the dataset for benchmarking: A $(2,2)$, B $(3,5)$, C $(6,7)$, and D $(9,8)$. Model (3) would first construct a CRS frontier (the solid line in Figure 9) by connecting the origin with any of A, B, C, or D that gives the largest slope. The dashed line is a virtual horizontal line for illustration purpose, which starts from building 0 and intersects CRS frontier and output axis at Y and X, respectively. The virtual efficient building, Y, can then be found by scaling down the actual efficient building, B, along the

CRS frontier. Finally, it compares the performance of Building *Y* with building 0, given that they have the same level of output, and calculates the efficiency score of building 0 as

$$\theta_0 = \frac{XY}{X0} \tag{4}$$

Model (4) is known as the CRS model as the frontier is constructed using only one single line, depicting any change of the input would change the output proportionally.

4.4.2 Variable Return to Scale

Based on the CRS DEA model, a variable return to scale (VRS) DEA model is proposed by Banker et al. (1984) as

$$\begin{aligned}
&\min_{\lambda_j} \theta_0 \\
&s.t. \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_0 x_{i0}; \quad i = 1,\ldots,m, \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}; \quad r = 1,\ldots,s, \\
&\sum_{j=1}^{n} \lambda_j = 1; \\
&\lambda_j \ge 0; \qquad j = 1,\ldots,n
\end{aligned} \tag{5}$$

As shown in the above equation, the VRS model has one additional constraint that the summation of all the lambdas is equal to 1 when compared to CRS model. This additional constraint limits the search of virtual efficient target to be the convex combination of efficient DMU's on the frontier, rather than scaling up or down any individual efficient DMU.

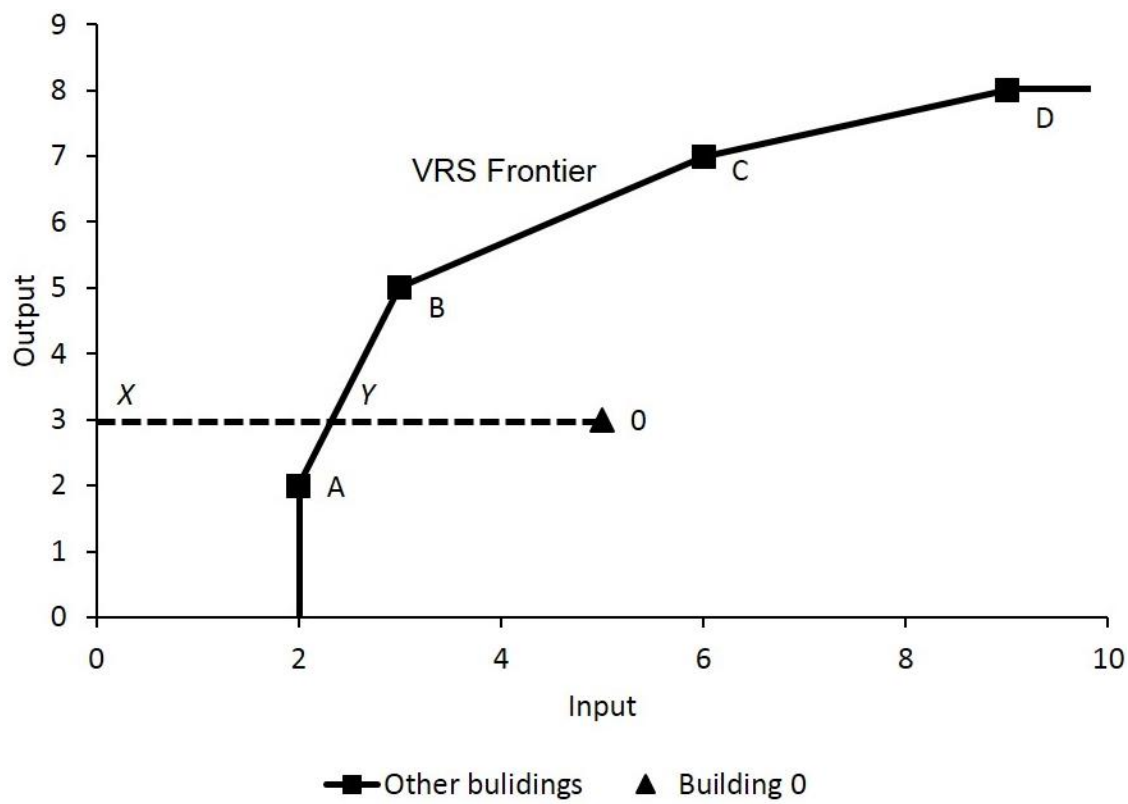

Figure 10 – Illustration of VRS DEA model

Figure 10 graphically illustrates how Model (5) works in a simplified one-input-one-output scenario. Assume we have the same buildings in Figure 9: building 0 (5,3), A (2,2), B (3,5), C (6,7), and D (9,8). Model (5) would first construct a VRS frontier by connecting all efficient DMU's (A, B, C, and D in this case). Again, the dashed line is a virtual horizontal line for illustration purpose, which starts from building 0 and intersects

VRS frontier and output axis at Y and X, respectively. The virtual efficient building, Y, can then be found by calculating the convex combination of building A and building B. Finally, it compares the performance of building Y with building 0, given that they have the same level of output, and calculates the efficiency score using Equation (4).

4.4.3 Non-Controllable DEA Model

DEA models benchmark performance based on constructing efficient DMU by the scaling or convex combination of existing DMUs. That scaling or convex combination, however, would make little sense if we have non-controllable variables such as weather conditions in the model, because they cannot be varied at the discretion of either property managers or tenants. However, those weather conditions need also be taken into consideration because they would make an impact on building energy consumption.

Based on the previous CRS and VRS models, a non-controllable variable (NCN) model can be further expressed as (Cooper et al., 2006):

$$
\begin{aligned}
&\min_{\lambda_j} \theta_0 \\
&s.t. \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_0 x_{i0}; \quad x_{i\bullet} \in X^C \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}; \qquad y_{r\bullet} \in Y^C \\
&\sum_{j=1}^{n} x_{ij}\lambda_j = \theta_0 x_{i0}; \quad x_{i\bullet} \in X^N \\
&\sum_{j=1}^{n} y_{rj}\lambda_j = y_{r0}; \qquad y_{r\bullet} \in Y^N \\
&L \le \sum_{j=1}^{n} \lambda_j \le U; \\
&\lambda_j \ge 0; \qquad\qquad j = 1,\ldots,n
\end{aligned}
\tag{6}
$$

where $x_{i\bullet}$ is the ith input variable of all DMUs; $y_{\mathrm{r}\bullet}$ is the rth output variable of all DMU's; X^C is the set of all controllable input variables; X^N is the set of all non-controllable input variables; Y^C is the set of all controllable output variables; and Y^N is the set of all non-controllable output variables. L and U set the lower bound and upper bound of the summation of weights, respectively. If $L = 0$ and $U = +\infty$, then we have a CRS model; if $L = U = 1$, then we have a VRS model. Model (6) essentially treats controllable variables, both input and output variables, and non-controllable variables differently by making controllable variables scalable and non-controllable variables constant.

4.5 Outlier Detection

Outliers are outlying observations that appear to deviate markedly from other observations of the sample in which they occur (Grubbs, F.E., 1969). One of the concerns about using non-parametric models, such as DEA is the existence of outliers (Khezrimotlagh, 2013), because they may dramatically change the shape of DEA efficient frontiers and give misleading efficiency scores to other non-efficient DMUs. The goal of identifying and removing outliers is to make the remaining DMUs more comparable and therefore, the efficiency scores more meaningful, as DEA is a peer-to-peer comparison method. There are two types of outliers that will be detected and removed: super-efficient and super-inefficient DMUs.

Previously, Lu et al. (2014) and Wang et al. (2015) utilized a data cloud analysis method to identify outliers for DEA model to benchmark building energy. The method takes one variable at a time, and iteratively calculates the log ratio of data volume change when one or more observations are removed. The approach identifies the outlying observations based on the log ratio of volume change. This method, however, is subject to two main limitations: (a) It takes only one variable each time, which limits one of the most appealing advantages of DEA method that it can simultaneously consider multiple input and output variables (Tran et al., 2010); and (b) The identified observations are geographically (spatially) outlying observations but they may not necessarily be the outliers in the context of the DEA formulation as outliers in the DEA method are simply those super-efficient or super-inefficient DMUs.

An effective and easy method to detect super-efficient outliers is suggested by Tran et al. (2010). Recall that parameter λ_j in CRS model (Model (3)), VRS model (Model (5)), and NCN model (Model (6)) represents the weight assigned to the jth DMU to construct the virtually efficient DMU for evaluating DMU_0. To find the efficiency scores of all n DMUs, the corresponding model needs to be solved n times, generating an $n \times n$ matrix, M_λ, containing all the λ's as follows

$$M_\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nn} \end{bmatrix} \tag{7}$$

The ith row and jth column of M_λ represents the weight assigned to the jth DMU to construct the virtually efficient DMU for evaluating the ith DMU. Outliers that perform significantly better than other DMUs are the ones that would always be selected to construct the virtually efficient DMU. They can therefore be identified through the number of occurrences during the construction of virtually efficient DMU as

$$C_j = \sum_{i=1}^{n} 1\{\lambda_{ij} > 0\} \tag{8}$$

Where $1\{\lambda_{ij} > 0\}$ is an indicator function and it returns 1 if $\lambda_{ij} > 0$ is true; otherwise 0. The outliers can also be identified through the cumulative weight during the construction of virtually efficient DMU as

$$S_j = \sum_{i=1}^{n} \lambda_{ij} \tag{9}$$

Outliers are DMUs performing significantly better than their peers, so they are the ones with surprisingly high number of occurrences and value of cumulative weight. After each time running the model and calculating the corresponding C_j and S_j $(j = 1, \dots, n)$, the DMU with C_j and S_j higher than certain thresholds would be identified as the outlier and be removed from the dataset. Although the thresholds can be subjective and were not discussed in the literature proposing this method, potential thresholds will be suggested in a later chapter. The process stops once a desired degree of convergence in the weights has been reached (Tran et al., 2010).

After those super-efficient DMUs are identified and removed, significant increases of the efficiency scores of most DMUs are expected. We can re-run the DEA model based on the rest of data and try to identify the other type of outliers, those super-inefficient DMUs. This can be done by checking the efficiency scores of all DMUs in the rest of data, and those very low scores, for example 0.2 or less, are suspicious and can be identified as super-inefficient outliers (Cooper et al., 2011).

4.6 Efficiency Analysis

Three different efficiencies can be generally produced and analyzed via DEA model when benchmarking building energy consumption: overall efficiency, pure technical efficiency, and scale efficiency (Chauhan et al., 2006; Lee, 2008; Lee and Lee, 2009; Wang et al., 2015). Figure 11 illustrates how each efficiency is calculated using the same example shown in Figure 9 and Figure 10: building 0 $(5,3)$, A $(2,2)$, B $(3,5)$, C $(6,7)$, and D $(9,8)$.

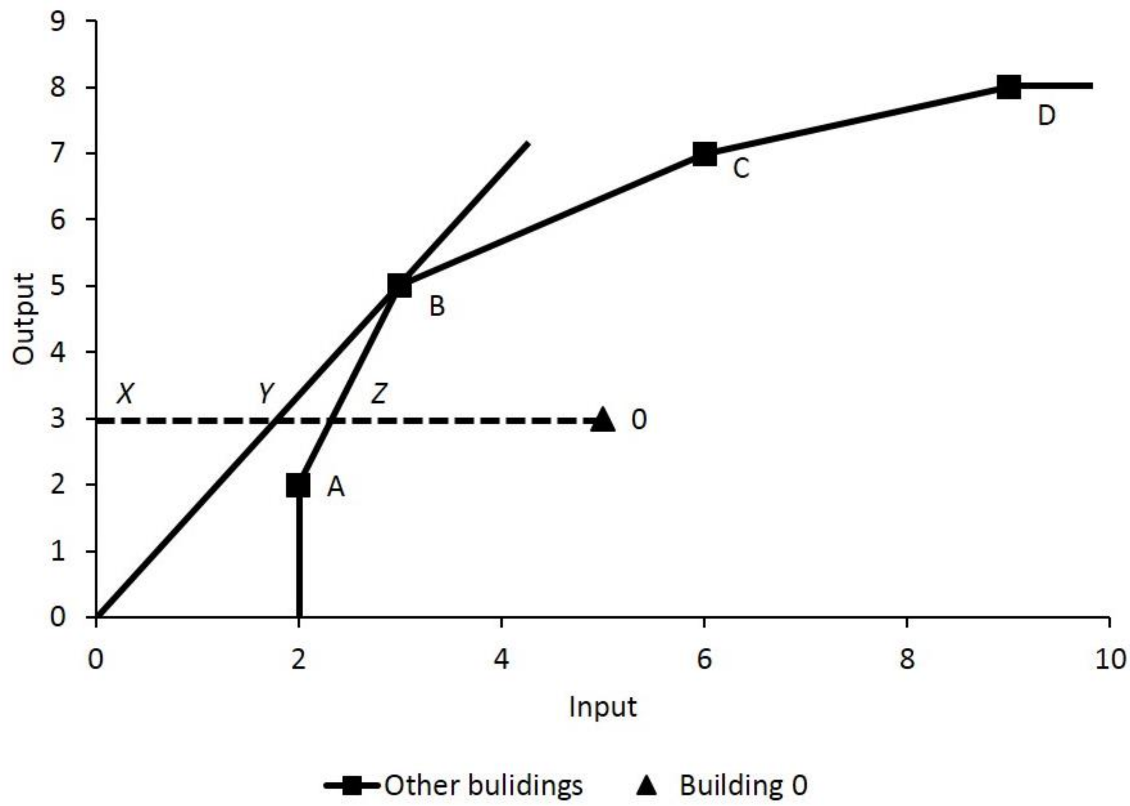

Figure 11 – Illustration of different efficiency scores in DEA model

In Figure 11, two solid lines represents the CRS efficient frontier and VRS efficient frontier, respectively, which is the same as shown in Figure 9 and Figure 10. Again, the dashed line is a virtual horizontal line for illustration purpose, which starts from building 0 and intersects VRS frontier, CRS frontier, and output axis at Z, Y, and X, respectively. All three types of efficiencies can then be calculated as follows (Chauhan et al., 2006; Lee, 2008; Lee and Lee, 2009; Wang et al., 2015):

$$\begin{aligned} &\text{Overall efficiency } = XY/X0 \\ &\text{Scale efficiency } = XY/XZ \\ &\text{Pure technical efficiency } = XZ/X0 \end{aligned} \tag{10}$$

Overall efficiency is a measure by which DMUs are evaluated for their performance relative to the best DMU in the comparison pool, and its value is influenced by scale efficiency (Chauhan et al., 2006). Scale efficiency represents the efficiency level in terms of scale of economics. Pure technical efficiency is the efficiency that has the scale influence removed, and it generally represents the efficiency level of management practices. Each of the above three efficiency scores has a range of 0 to 100%, and the higher the efficiency score, the more efficient the performance. A DMU receiving 100% efficiency score is an efficient DMU, and one receiving less than 100% is inefficient.

The relationship among the three efficiencies is given as (Chauhan et al., 2006; Lee, 2008; Lee and Lee, 2009):

$$\text{Overall efficiency} = [\text{Pure technical efficiency}] \times [\text{Scale efficiency}] \quad (11)$$

In the context of building energy benchmarking, pure technical efficiency represents the goodness of management practices. Namely, higher pure technical efficiency score means that the facility manager is managing the property in a more energy-efficient manner. Scale efficiency represents the level of energy efficiency in terms of the scale of the building, such as the floor area of the building and number of bedrooms in a building.

In fact, the efficiency score calculated from the CRS model is overall efficiency, and the one calculated from the VRS model is pure technical efficiency. Scale efficiency can therefore be derived as overall efficiency divided by pure technical efficiency.

4.7 Window Analysis

Input variables of DMUs in this study are the energy consumptions of buildings over multiple time periods (months), and it is likely that the performance of a particular building varies a lot from time to time. Window analysis is therefore often suggested to measure the sensitivity and stability of efficiency scores of DMUs when dealing with time series data.

Let n be the number of DMUs to be analyzed, T be the total number of time periods, and k ($k \leq T$) be the window size, i.e. the number of periods in each window. The number of windows, ω, can be therefore calculated as $\omega = T - k + 1$, and the number of DMUs in each window can be calculated as nk. Table 5 shows an example of creating window analysis on DMUs with $T = 12$ periods (months) from January 2009 to December 2009, $k = 3$, and therefore $\omega = 10$.

Table 5 – Periods corresponding to each window in DEA window analysis

Windows	Periods Corresponding to Each Window											
Window 1	Jan-09	Feb-09	Mar-09									
Window 2		Feb-09	Mar-09	Apr-09								
Window 3			Mar-09	Apr-09	May-09							
Window 4				Apr-09	May-09	Jun-09						
Window 5					May-09	Jun-09	Jul-09					
Window 6						Jun-09	Jul-09	Aug-09				
Window 7							Jul-09	Aug-09	Sept-09			
Window 8								Aug-09	Sept-09	Oct-09		
Window 9									Sept-09	Oct-09	Nov-09	
Window 10										Oct-09	Nov-09	Dec-09

For each window, DMUs are not only compared with other DMUs in the same window, but also with DMUs from other periods. For instance, in Window 1, DMU 1 in period of Jan-09 is not only compared with the other $n-1$ DMUs in period of Jan-09, but also compared with itself and the other $n-1$ DMUs in periods of Feb-09 and Mar-09. With that being said, the model would consider there are nk DMUs in each window and treat the same DMU from different periods as different DMUs. In window analysis, a DMU that is efficient in most periods, regardless of the window, is likely to be truly efficient relative to other DMUs. On the contrast, a DMU that is only efficient in certain periods of certain windows may be efficient because of external circumstances (Yue, 1992).

Results of window analysis can be presented in the format shown in Table 5, with each period replaced by the efficiency score for that period. In Table 5, each period from Mar-09 to Oct-09 would have three efficiency scores calculated from three different windows, and the average of those three efficiency scores can be taken to reflect the level of efficiency for that period. Several other statistics, such as average, standard deviation, range, etc. across the entire period (twelve months in this case) can also be calculated to reflect the sensitivity and stability of efficiency scores of DMUs.

An important parameter in window analysis is the determination of the window size. If the window size is too small, there may not be enough DMUs analyzed in the current window period and thus not enough discrimination in the results. On the other hand, if the window size is too large, DMUs may become not comparable because significant changes may have occurred in a wide range of period (Cooper et al., 2011). Unfortunately, there is no current theory supports the determination of window size (Cullinane et al., 2004), and a commonly utilized window size is three periods (Yue, 1992; Cullinane et al., 2004). The same window size of three periods is also chosen for this research.

CHAPTER 5. RESULSTS AND INTERPRETATION

5.1 Dataset and Variable Selection

The proposed approach is applied to a dataset provided by a utility management and energy service company in the multifamily housing industry. It contains the information about both building characteristics and energy consumption of 124 low-rise (1-4 floors) multifamily properties in 15 different states in the United States, such as Georgia, North Carolina, and Virginia.

Based on the literature review on variable selection for all DEA models applied to different building types when benchmarking energy efficiency, it is found that two output variables are commonly utilized in DEA models for building energy benchmarking: total floor area and the number of occupants. Those two variables would also be selected for this research. According to consultation with industrial experts and data provider, four additional output variables are also selected to be included in this research: number of apartments, number of bedrooms, number of washing machines, and number of parking spaces.

Of course, the input variable, energy consumption, should also be included. Energy consumption data includes 12 monthly electricity usage of each property from January 2009 to December 2009. One thing to notice is that this research would not utilize EUI, which is commonly used in previous research (Önüt and Soner, 2006; Lee and Lee, 2009; Lu et al., 2014; and Wang et al., 2015), as an input variable due to the fact that including any ratio variable may violate the convexity assumption of DEA model as discussed in

Section 4.1. This research would rather treat monthly energy consumption as input variable and floor area as output variable. Table 6 summarizes major statistics of input and output variables to be included in this research.

Table 6 – Summaries of major statistics of input and output variables

	Minimum	Mean	Std. Dev.	Maximum
Input variable				
Monthly energy (Kwh)	370	68,002	228,278	3,787,374
Output variable				
Total floor area (SF)	42,850	376,369	252,070	1,699,453
# of residents	83	706	737	5,500
# of apartments	80	335	214	2,346
# of bedrooms	1	541	309	2,530
# of parking spaces	25	560	300	1,872
# of washing machines	1	261	155	936

As Table 6 suggests, there is a wide range of values for both input and output variables. Particularly, there are instances in the dataset that do not make any sense and cannot be directly used by DEA model for building energy benchmarking. For example, a property with 386 residents inside has only one bedroom. Further steps of remediating data errors are therefore needed and will be conducted in the next section.

5.2 Data Errors Remediation

To give more detailed information about how many DMUs are good or problematic, and further how many problematic DMUs can be remediated, we classify all DMUs in the dataset into four categories based on the discussion in Section 4.1: good DMUs, DMUs with input data errors, DMUs with only one output data error, and DMUs with more than one output data errors. Table 7 outlays the number of DMUs in each of those four categories before remediating data errors. Recall that DMUs in category three, i.e. with only one output data error, are the ones we are trying to remediate.

Table 7 – Number of DMUs in each category before remediating data errors

	Number of DMUs
Total DMUs	124
Good DMUs	89
DMUs with input data errors	15
DMUs with only one output data error	19
Data error with # of washing machines	18
Data error with # of bedrooms	1
DMUs with more than one output data errors	1

As shown in Table 7, DMUs with only one output data error either have error in the number of washing machine or the number of bedrooms. The regression model that treats the number of washing machines as the dependent variable and all other output variables as independent variables using good DMUs is:

$$NWM = -23.74 + 1.00 \times NA \tag{12}$$

where NWM is the number of washing machines, and NA is the number of apartments. Other output variables are not selected in the regression model because they are not significant (i.e. p-values are > 0.05). The p-value of NA in this model is < 0.001, and $R^2 = 0.84$.

The regression model that treats the number of bedrooms as the dependent variable and all other output variables as independent variables using good DMUs is:

$$NB = -45.54 + 1.27 \times NA + 0.16 \times NPS + 0.13 \times NR \qquad (13)$$

where NB is the number of bedrooms, NA is the number of apartments, NP is the number of parking spaces, and NR is the number of residents. Other output variables are not selected in the regression model because they are not significant (i.e. p-values are > 0.05). The p-value of NA in this model is < 0.001, of NPS in this model is 0.003, of NR in this model is < 0.001, and $R^2 = 0.89$.

Because R^2 values are high and independent variables are significant in both linear regression models, they can be utilized to remediate data errors. Table 8 outlays the number of DMUs in each of the four categories after remediating data errors.

Table 8 – Number of DMUs in each category after remediating data errors

	Number of DMUs
Total DMUs	124
Good DMUs	108
DMUs with input data errors	15
DMUs with only one output data error	0
DMUs with more than one output data errors	1

The process of data error remediation is in fact the process of building linear regression model and using the built model for prediction, which is computationally efficient and can be done in polynomial time.

5.3 Model Formulation

As shown in Figure 11, properties in the dataset for this research are from different states in the U.S., so the weather influence on building energy consumption needs to be taken into consideration as discussed in Section 4.3. Monthly TDD data of each property is collected from weather data depot (WDD, 2016) by specifying the zip code for each property and fixing the balance point temperature at 65^{o}F.

The complete data set for DEA model of this research has eight variables in total, including one input variable and seven output variables. Table 9 summarizes major statistics of all variables to be included in the DEA model for this research.

Table 9 – Summaries of major statistics of input and output variables of this research

	Minimum	Mean	Std. Dev.	Maximum
Input variable				
Monthly energy (Kwh)	370	71,546	238,188	3,787,374
Output variable				
Total floor area (SF)	42,850	339,653	148,359	900,000
# of residents	83	704	662	4,477
# of apartments	80	314	123	936
# of bedrooms	118	531	268	2,520
# of parking spaces	25	563	299	1,872
# of washing machines	25	292	133	936
Monthly TDD	93	403	204	1,547

The reason energy consumption is modeled as an input variable is based on the principal of DEA modeling that variables need to be minimized are input variables and variable need to be maximized are output variables. By using DEA model for building energy benchmarking, we are either trying to minimize the building energy consumption while maintaining the same level of scale (such as floor area), or we are trying to maximize the scale while maintaining the same level of energy consumption.

Both CRS NCN DEA and VRS NCN DEA models are utilized in this research to calculate the corresponding efficiency scores of each DMU in each period as discussed in Section 4.4 and Section 4.6. Monthly TDD would be treated as an non-controllable output variable, or Y^N in model (6), and all other variables would be treated as controllable input or output variables.

Both CRS NCN DEA and VRS NCN DEA models are linear programming problems, and they are modeled in Excel and solved by Excel Solver in this research. The Excel Solver solves linear programming problems using Simplex algorithm, which is a fast and efficient algorithm for solving linear programming problems. The model should iteratively choose one property as the target DMU at a time, and should run through the entire period (twelve months in this research).

5.4 Outlier Detection

Based on the discussion in Section 4.5, super-efficient outliers are DMUs with large number of occurrences and high cumulative weight when constructing the virtually efficient DMU, as shown in Equation (8) and Equation (9), respectively. M_{λ}, as shown in Equation (7), would be calculated after running the DEA model. Table 10 summarizes the number of occurrences and cumulative weight of potential outlying DMUs, those that are referenced at least once to construct the virtually efficient DMU.

To detect real outliers, thresholds for number of occurrences and cumulative weight need to be set up. DMUs having both numbers higher than the thresholds are identified as outliers. Although Tran et al. (2010) did not specify the thresholds in their research, I would suggest utilize median plus 2x standard deviation as the threshold. Any DMU with both number of occurrences and cumulative weight higher than median plus 2x standard deviation can be considered as significantly larger than the vast majority, and can therefore be identified as an outlier. Notice that the threshold criteria are not unique, and I am just suggesting one possible solution. The reason I chose median instead of average is that any extremely large number can increase the average significantly.

Table 10 – Summary of potential outlying DMUs for January 2009

DMU	Number of occurrences	Cumulative weight
3	61	20.97
5	46	8.80
18	11	2.34
30	5	1.59
59	4	2.40
63	8	3.55
66	1	1.00
67	11	4.15
69	2	1.68
70	20	3.74
73	29	9.45
75	1	1.00
79	2	1.29
80	3	1.72
102	88	41.78
104	2	1.58

Table 11 – Threshold criteria to detect outliers for January 2009

	Number of occurrences	Cumulative weight
Median	6.50	2.37
Std. Dev.	25.55	10.66
Median + 2x Std. Dev.	57.59	23.69

Table 11 summarizes the threshold for number of occurrences and cumulative weight to detect outliers in January 2009. Based on those thresholds, one DMU in Table 10 can be identified as outliers: DMU 102, which has both the number of occurrences and cumulative weight higher their respective thresholds. Recall that energy consumption in this research is a time series, and has data from January 2009 to December 2009 for twelve months. Outliers therefore need to be removed every month. The process should be repeated for every month, and each time should incorporate the energy consumption and TDD of that particular month for outlier identification and removal.

After repeating the outlier detection process twelve times, 15 outliers were detected and removed from the original 108 DMUs, leaving 93 DMUs for further analysis. Recall our discussion in Section 4.5 that the above process identifies super-efficient outliers only, and the other type of outliers, the super-inefficient outliers, can be identified by choosing DMUs with efficiency score smaller than 0.2 (Cooper et al., 2011). So 76 DMUs are eventually left for comparison and efficiency score analysis.

Figure 12 shows the efficiency scores of all 108 DMUs before outliers are detected and removed in January 2009. As the figure shows, around 1/4 of DMUs are scored under 10%, and around 1/3 DMUs are scored under 20%. Those scores do not really mean that half of the properties did a very bad job in terms of building energy consumption, but it simply indicates that some super-efficient properties are in the dataset and are utilized as benchmark to measure the performance of other properties. This is not supposed to be the case as DEA is a peer-to-peer comparison tool, and DMUs included in the DEA model need to be comparable, namely no DMU is super-efficient or super-inefficient, or the resulting efficiency scores are otherwise misleading and useless.

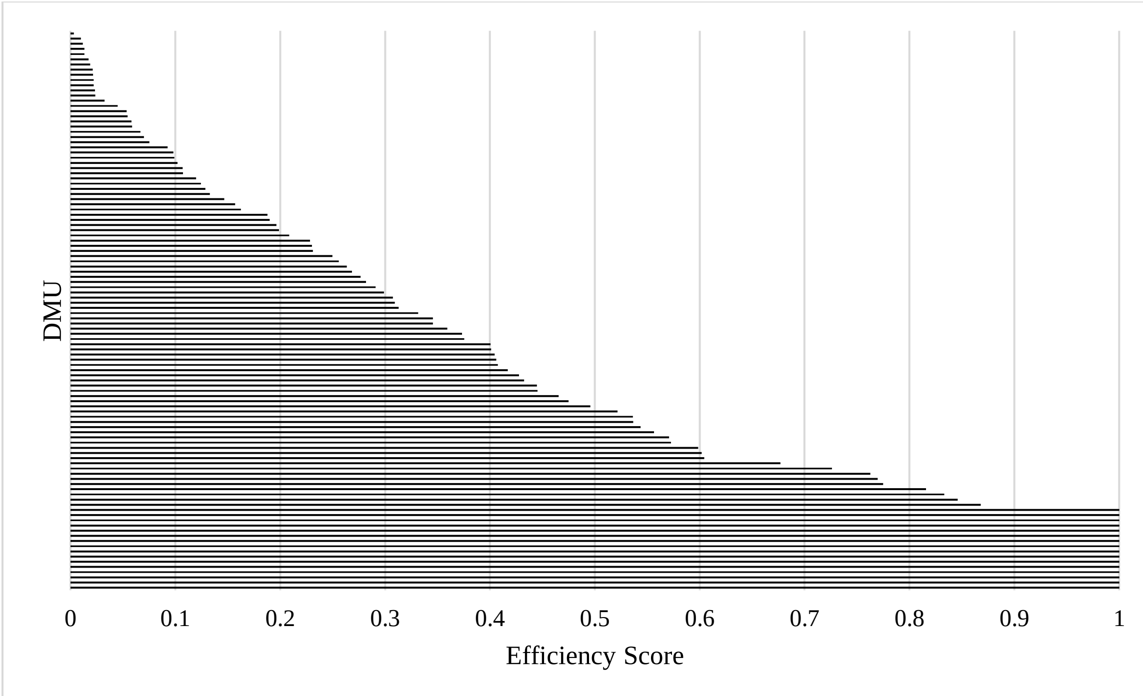

Figure 12 – Efficiency scores of all DMUs before outlier detection in January 2009

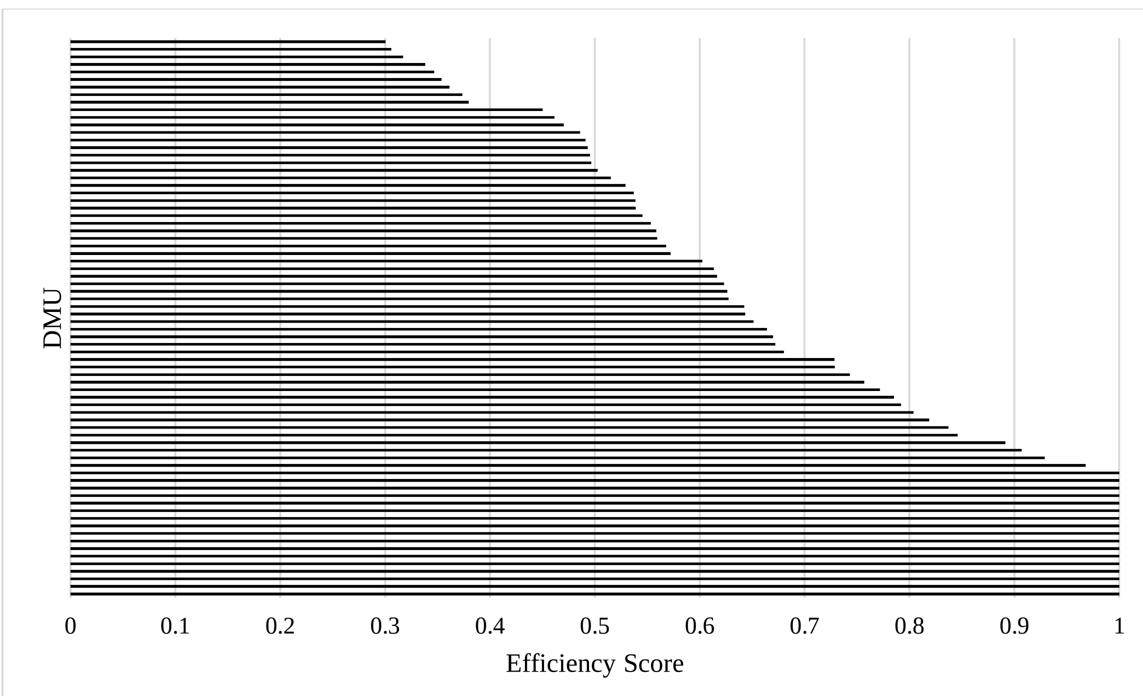

Figure 13 – Efficiency scores of all DMUs after outlier detection in January 2009

Figure 13 shows the efficiency scores of all 78 DMUs after outliers are detected and removed in January 2009. As the figure shows, no DMU receives a score below 20%, and more than 80% DMUs receive scores above 40%. The vast majority DMUs receive scores between 40% and 90%. One thing to notice is that it is important to do sanity check and make sure that the size of the new data set after removing outliers is large enough based on the three rules of thumb discussed in Section 4.2. Notice that only 32 DMUs are removed as outliers throughout the entire outlier detection process, but at least 39 DMUs (around 1/3 of the number of DMUs in Figure 12) would have otherwise been removed because of their super low efficiency score (under 20%) if we did not conduct outlier detection.

Recalling the three rules of thumb when checking sample size, it compares the number of DMUs in the dataset with the number of input and output variables. There are 76 DMUs in the final dataset, one input variable, and seven output variables. The sample size is greater than twice the product of the number of inputs and output, and it is also greater than three times the sum of the number of inputs and outputs. By further checking Figure 13, it is clear that the number of efficient DMUs does not exceed one third of the sample size. So the final dataset after outlier detection and removal satisfies the conditions previously mentioned and it can be used for further analysis.

The outlier detection and removal is based on the results of model formulation as discussed in Section 5.3, and the additional computational load is the calculation of both thresholds and identification of outliers, both of which can be done in polynomial time. The process identifies and removes DMUs that are either super-efficient or super-inefficient, making the rest of DMUs more comparable and appropriate for peer-to-peer comparison, as suggested in Figure 12 and Figure 13. Without out the outlier removal process, some extreme outcomes may be generated in the later efficiency analysis section, such as an outcome with only a very small portion of DMUs reaching 100% efficiency, while the vast majority of DMUs are scored under 20% efficiency.

5.5 Efficiency Analysis

According to the discussion in Section 4.6, three types of efficiency scores can be generated from the DEA model and analyzed: overall efficiency, pure technical efficiency, and scale efficiency.

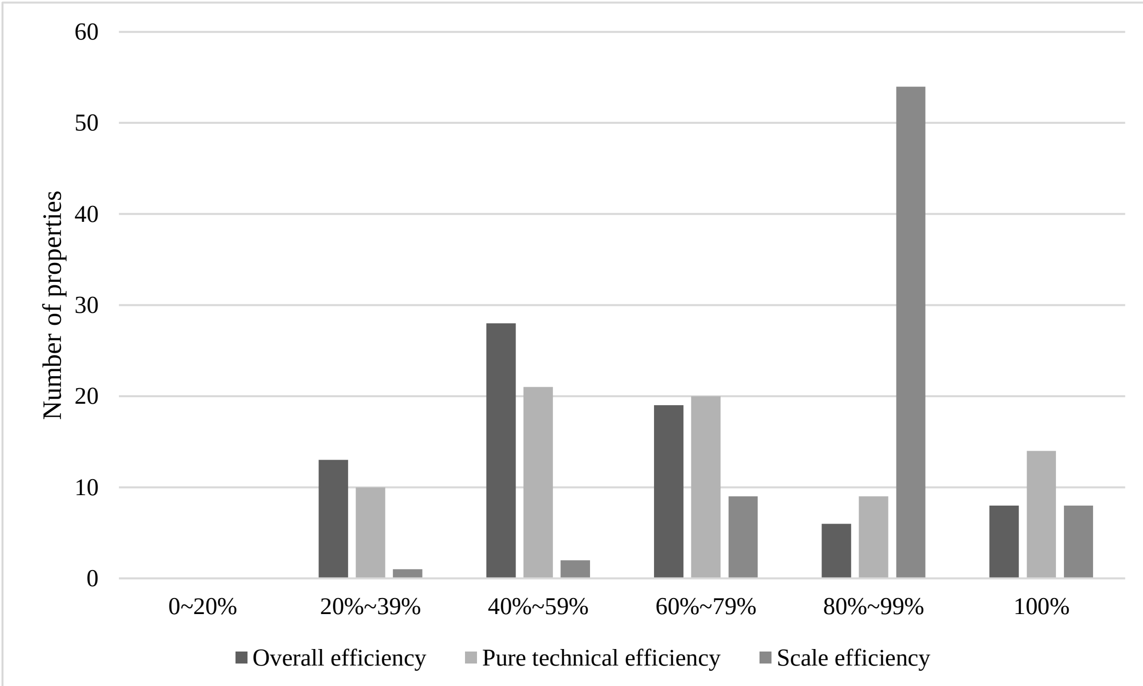

Figure 14 – Distribution of three efficiency scores for all properties in January 2009

Figure 14 shows the distribution of three efficiency scores for all properties in January 2009. Around 25% of properties have a pure technical efficiency score higher than 80%, and more than 45% of properties have a pure technical efficiency score higher than 60%. From the scale efficiency perspective, around 60% of properties have a scale efficiency score higher than 80%, and around 70% of properties have a scale efficiency score higher than 60%.

The distribution of three efficiency scores from February 2009 to December 2009 are also plotted and presented in Appendix A. Similar conclusions can be made from those of January 2009, which tells that the energy efficiency of properties under management are stable in general.

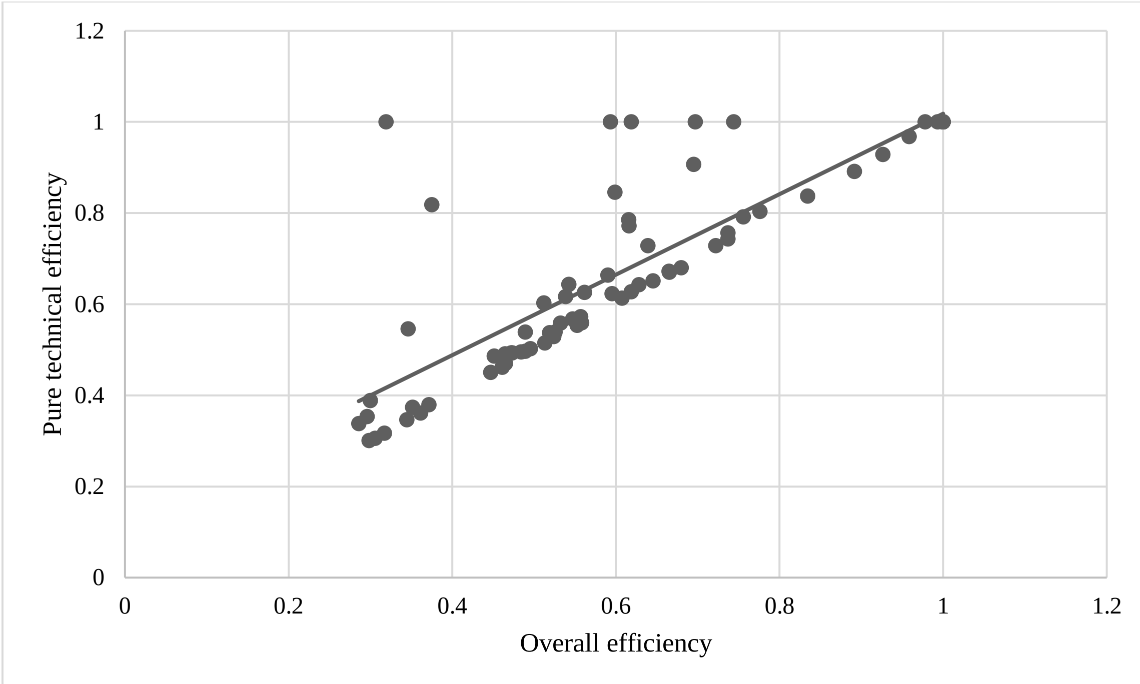

Figure 15 – Relationship between pure technical efficiency and overall efficiency in January 2009

Figure 15 reveals the relationship between pure technical efficiency and overall efficiency. It is clear to see that there is an up-trending relationship, and the correlation between them is 0.84, which means a higher pure technical efficiency would generally suggest a higher overall efficiency.

Figure 16 reveals the relationship between scale efficiency and overall efficiency. It can still be seen that there is an up-trending relationship, and the correlation between them is 0.30, which means a higher scale efficiency would generally suggest a higher overall efficiency.

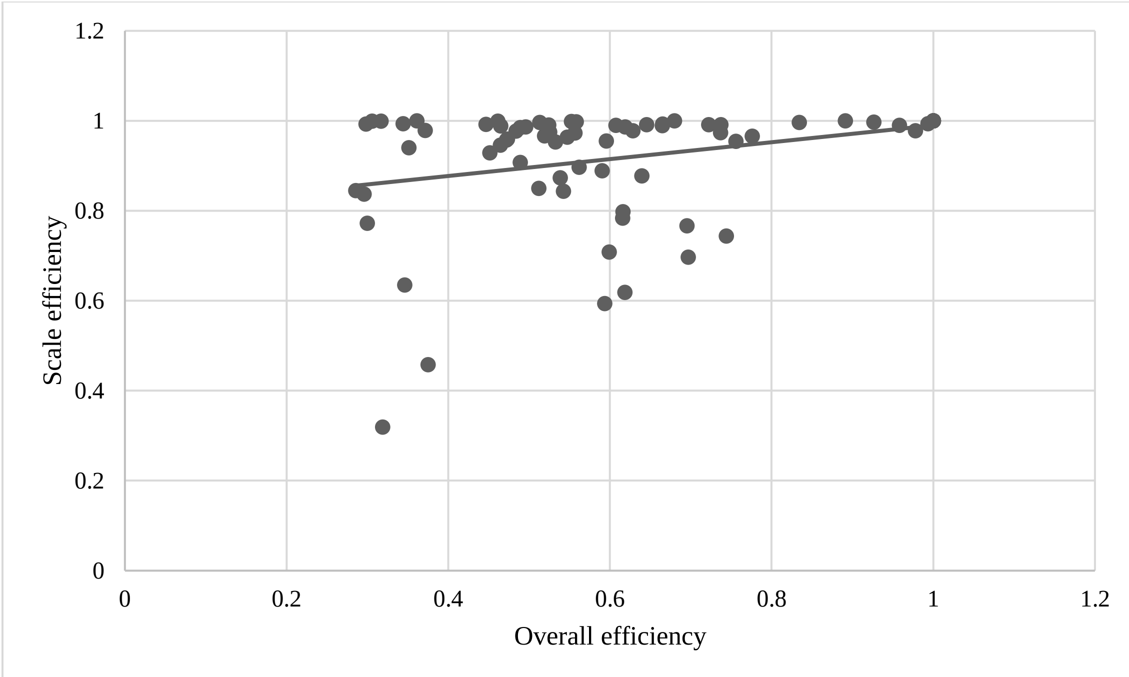

Figure 16 – Relationship between scale efficiency and overall efficiency in January 2009

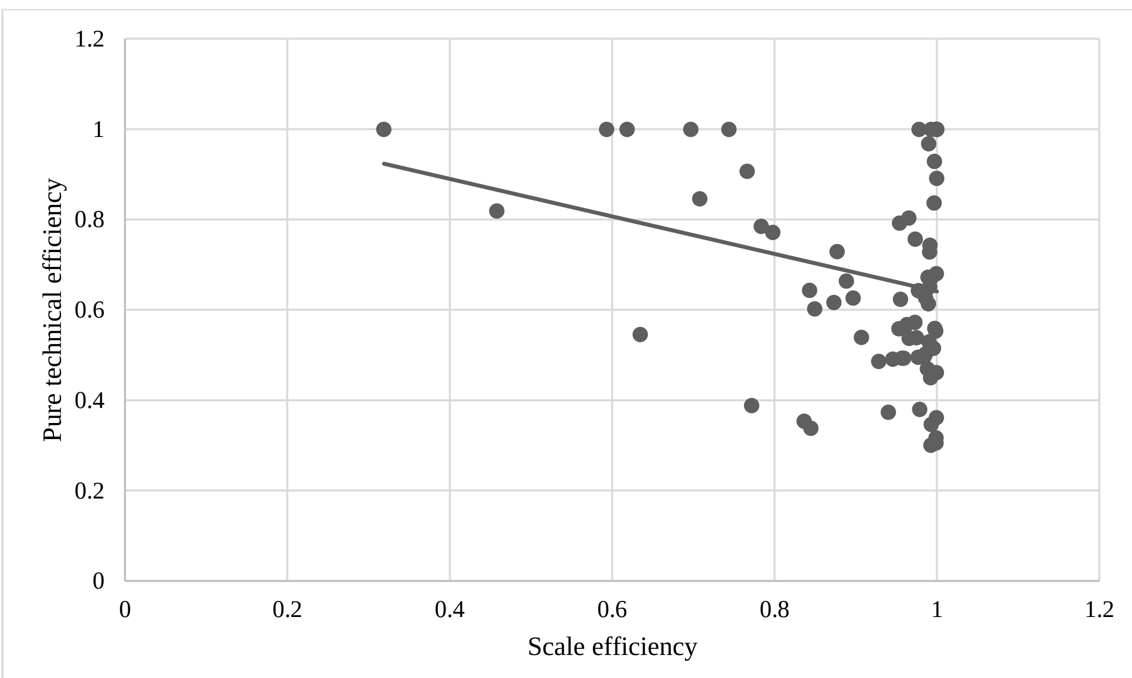

Figure 17 – Relationship between pure technical efficiency and scale efficiency in January 2009

Although both scale efficiency and pure technical efficiency would have an impact on overall efficiency, the correlation between pure technical efficiency and overall efficiency is much larger than that between scale efficiency and overall efficiency, which implies poor energy efficiency is largely attributed to poor energy management.

Figure 17 reveals the relationship between pure technical efficiency and scale efficiency. There is a slight downward trending relationship, and the correlation between them is -0.25, which means a higher scale efficiency would generally suggest a slightly lower pure technical efficiency. Additionally, the scattered plot also verifies one of the observation we had from Figure 14, that most buildings have a very high scale efficiency.

In fact, the relationships between those three efficiency scores found in this research are consistent with previous research (Lee and Lee, 2009).

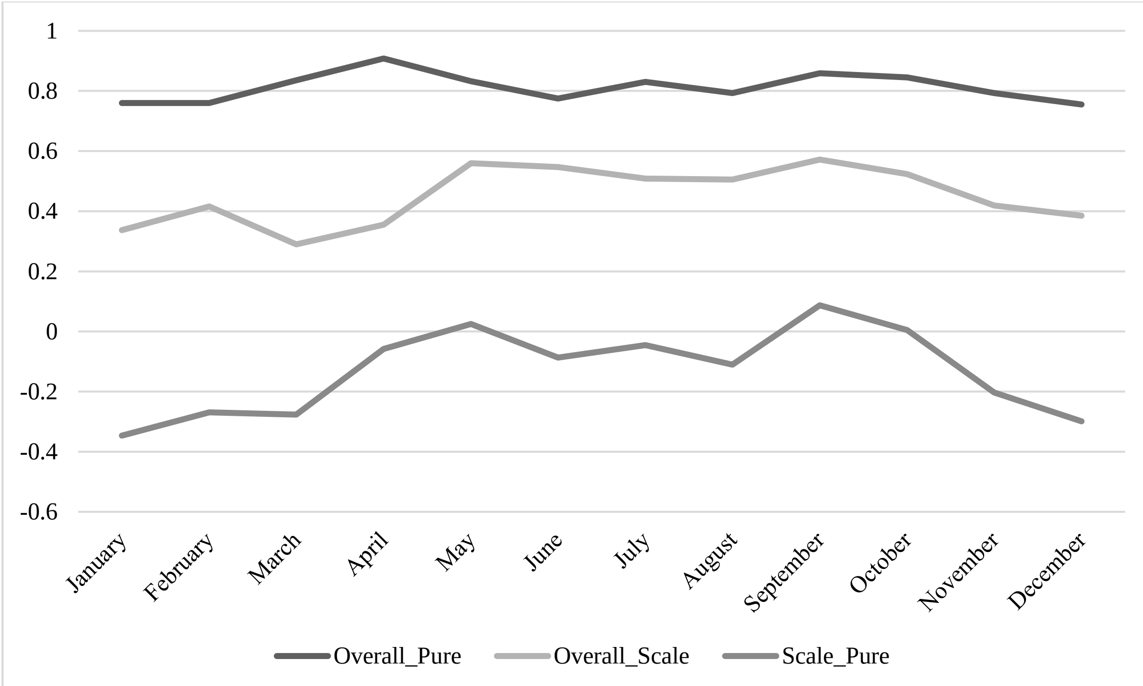

Figure 18 – Relationship between three efficiency scores from January 2009 to December 2009

Figure 18 reveals the relationship between all three efficiency scores from January to December. The correlation between overall efficiency and pure technical efficiency is around 0.8; the correlation between overall efficiency and scale efficiency is between 0.4 and 0.6; and the correlation between pure technical efficiency and scale efficiency varies between -0.3 to 0.1 and the correlation is negative for most time. The relationship makes sense in general because recall that overall efficiency is calculated by the product of pure technical efficiency and scale efficiency.

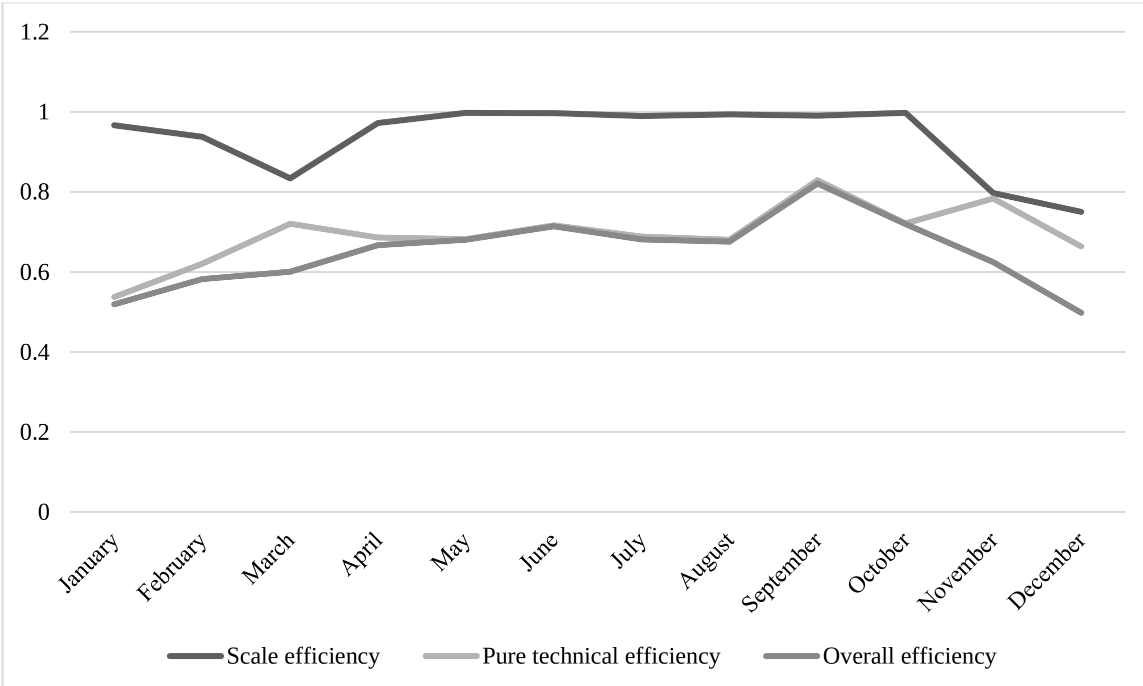

Figure 19 – Three efficiency scores of DMU 19 from January 2009 to December 2009

Figure 19 shows all three efficiency scores of DMU 19 from January 2009 to December 2009. It shows that DMU 19 had stable values for all three efficiency scores. The scale efficiency fell between 0.8 and 1, and both pure technical efficiency and overall efficiency fell between 0.6 and 0.8 for most time. The scale efficiency was relatively more stable across the entire period. It is interesting to find scale efficiency scores of most properties are relatively more stable than the other two efficiency scores, and they usually fall in between 80% and 100%. This actually makes intuitive sense because scale efficiency actually reflects the level of efficiency in terms of scale of building, which is generally supposed to be stable and unlikely to change from month to month.

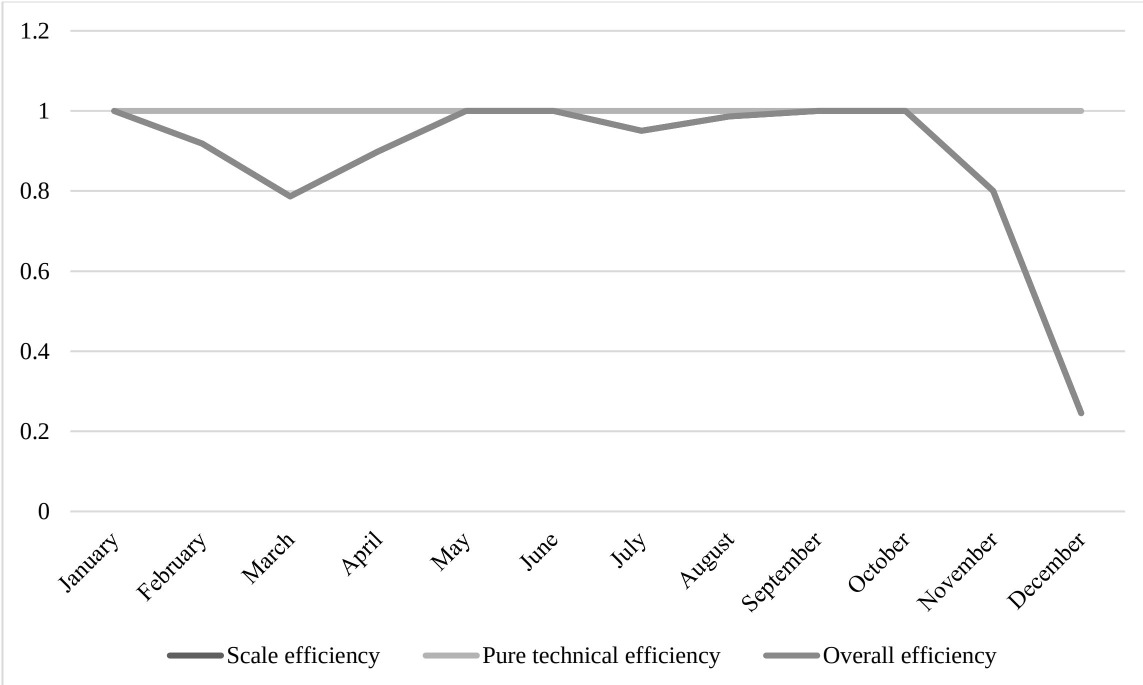

Figure 20 – Three efficiency scores of DMU 3 from January 2009 to December 2009

The knowledge of efficiency scores over time provides valuable insight about the energy performance of properties. For example, Figure 20 illustrates all three efficiency scores of DMU 3 from January 2009 to December 2009. It shows that DMU 3 generally performed very well across the entire period with pure technical efficiency at 100% across the entire period, which means that the management practice of DMU 3 in terms of energy consumption is excellent. However, its overall efficiency showed some volatility and dropped to around 20% in December, resulting a corresponding drop of its scale efficiency score, meaning that there are some other DMUs reaching a more significant performance improvement in the summer season.

As the DEA is a peer-to-peer benchmarking method, the performance of one DMU in the model is dependent on its relative performance compared with the efficient frontier. Adding one DMU to the existing data may or may not change the shape of the frontier. If the added DMU does not change the frontier shape, then it is safe to run the model as it is and there is no need to rebuild the model. However, if the added DMU changes the frontier shape, then the model needs to be rebuilt and the efficiency scores of all other DMUs may also change accordingly. One easy and quick way to check whether model rebuild is needed is to evaluate the added DMU using the existing model, and if the efficiency score of the added DMU is less than 100%, then model rebuilt is not needed.

5.6 Window Analysis

As previously discussed in Section 4.7, there is a deterministic linear relationship between the number of windows and the size of each window. However, there is no current theory supports the determination of window size (Cullinane et al., 2004), and a commonly utilized window size is three periods (Yue, 1992; Cullinane et al., 2004). The same window size of three periods is also chosen for this research and the results window analysis can be presented in a similar format shown in Table 5.

To be consistent with previous efficiency analysis and for better comparison, I selected the same DMUs for window analysis, and Table 12 and Table 13 show the results of window analysis for DMU 19 and DMU 3, respectively. One thing to notice is that the efficiency score in window analysis for a certain period should not exceed that in efficiency analysis for the same period. This is simply because the efficiency score can only decrease or remain the same by incorporating more DMUs into analysis.

Several statistics can be analyzed from window analysis. The average efficiency score of each property for each month and for the entire period can be calculated to show how well it performs in each month and for the entire year. Note that the moving average is not as good as simple average as it mixes efficiency scores from different months and does not show the performance at a given time. The standard deviation of efficiency scores can also be calculated, and it shows the volatility of performance of a particular property. Finally, the range of efficiency scores can be utilized to give an overview of how widely the scores are distributed.

The interpretation of scores in window analysis is essentially the same with that from efficiency analysis. For example, as shown in Table 13, DMU 3 was scored 51.5% in December 2009. It means that DMU 3 was 51.5% efficient among all other DMUs for the month of December 2009. The other way to understand can be that the most efficient property consumes only 51.5% of the energy DMU 3 consumed in December 2009, and its scale is at least as large as that of DMU 3.

Table 12 – Results of pure technical efficiency window analysis for DMU 19

Windows	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sept-09	Oct-09	Nov-09	Dec-09
Window 1	53.89%	53.73%	53.73%									
Window 2		52.39%	52.82%	68.72%								
Window 3			52.43%	67.44%	62.41%							
Window 4				67.66%	65.01%	65.41%						
Window 5					68.65%	65.49%	62.85%					
Window 6						72.57%	64.29%	66.61%				
Window 7							63.25%	66.11%	83.31%			
Window 8								66.97%	76.53%	71.80%		
Window 9									70.97%	67.89%	67.42%	
Window 10										69.03%	66.75%	63.73%
Average	53.89%	53.06%	52.99%	67.94%	65.36%	67.82%	63.46%	66.56%	76.94%	69.57%	67.09%	63.73%
Mean	SD	LDY	LDP									
65.00%	7.31%	12.34%	30.92%									

Notes: Mean (average score for the twelve month period)
SD (standard deviation for the period)
LDY (largest difference between scores in the same month)
LDP (largest difference between scores across the entire period)

Table 13 – Results of pure technical efficiency window analysis for DMU 3

Windows	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sept-09	Oct-09	Nov-09	Dec-09
Window 1	100%	100%	100%									
Window 2		100%	100%	100%								
Window 3			100%	100%	100%							
Window 4				100%	100%	79.46%						
Window 5					100%	88.43%	83.28%					
Window 6						100%	89.50%	94.67%				
Window 7							81.61%	93.04%	100%			
Window 8								93.96%	100%	100%		
Window 9									100%	98.75%	87.57%	
Window 10										100%	84.67%	51.50%
Average	100%	100%	100%	100%	100%	89.30%	84.80%	93.89%	100%	99.58%	86.12%	51.50%
Mean	SD	LDY	LDP									
94.21%	10.40%	20.54%	48.50%									
Notes:	Mean	(average score for the twelve month period)										
	SD	(standard deviation for the period)										
	LDY	(largest difference between scores in the same month)										
	LDP	(largest difference between scores across the entire period)										

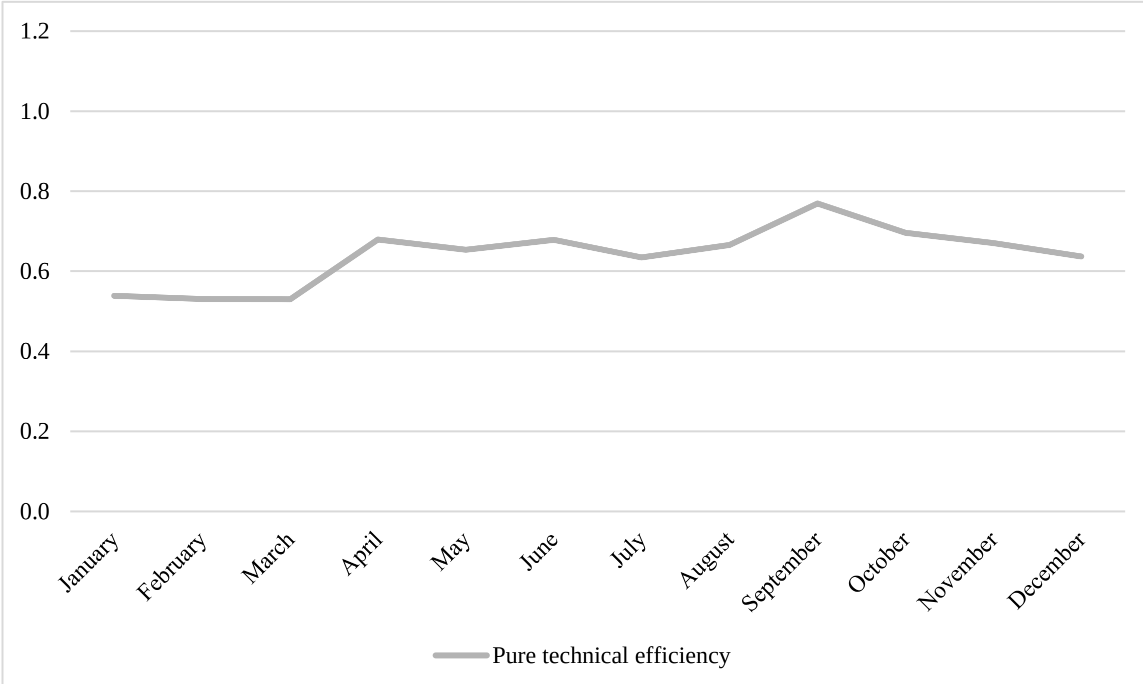

Figure 21 – Average pure technical efficiency score of DMU 19 from window analysis between January 2009 and December 2009

The monthly average pure technical efficiency score of DMU 19 from window analysis is plotted in Figure 21, and follows similar trend as shown in Figure 19. The efficiency score is relatively stable across the entire period. The mean efficiency score of DMU 19 is 65.00%, and its standard deviation is 7.31%. The largest difference in one month is 12.34%, which happened in September 2009. The largest difference across the entire period is 30.92%, which happened between February 2009 and September 2009. Those two differences mean that the efficiency score of each month is relatively stable, but the performance across the entire period is relatively volatile.

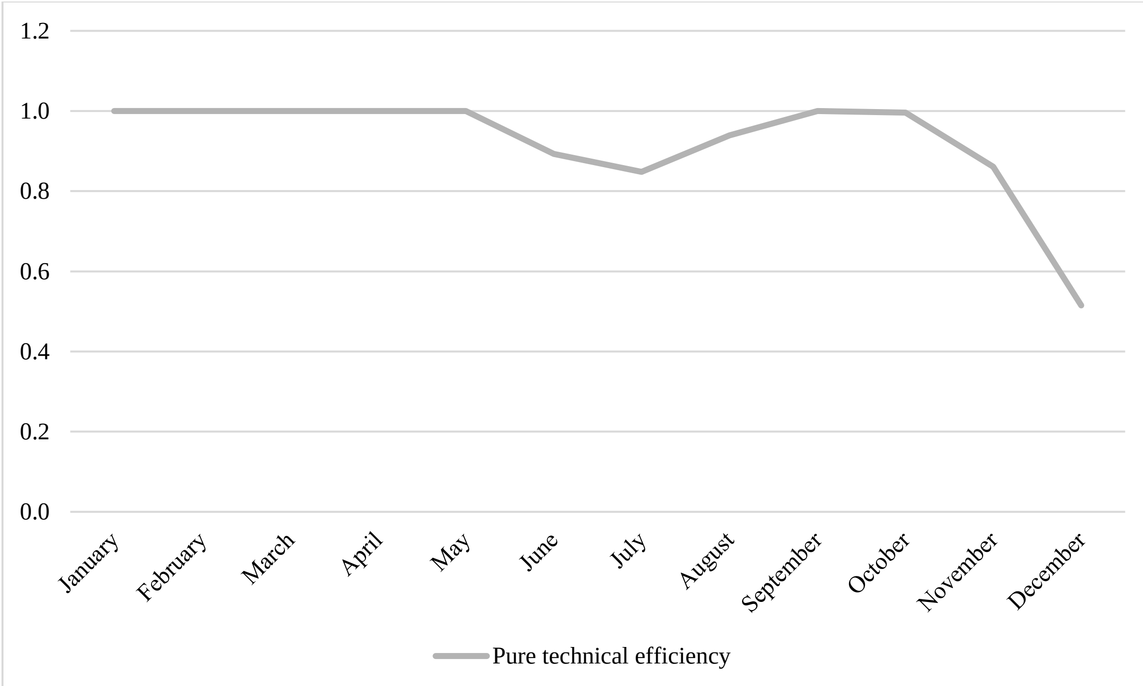

Figure 22 – Average pure technical efficiency score of DMU 3 from window analysis between January 2009 and December 2009

The monthly average pure technical efficiency score of DMU 3 from window analysis is plotted in Figure 22. Different from DMU 19, the window analysis results show some information different from the trend as shown in Figure 20. In Figure 20, the pure technical efficiency of DMU 3 is 100% across the entire period, which means the management practice was excellent all year long. Although the efficiency scores shown in window analysis are still pretty high, volatilities occur. Particularly in December 2009, the pure technical efficiency score dropped to around 50%. The mean efficiency score of DMU 3 is 94.21%, and its standard deviation is 10.40%. The largest difference in one month is 20.54%, which happened in June 2009. The largest difference across the entire period is 48.50%, which happened between the first five months of 2009 and December 2009. Compared with DMU 19, DMU 3 has a higher average efficiency score but a higher standard deviation as well. It represents that DMU 3 is generally more efficient than DMU 19, but December 2009 needs be to further examined to find why significant efficiency score drop happened in that month.

CHAPTER 6. CONCLUSIONS

This chapter intends to summarize and offer concluding remarks for this research. Particularly, it addresses the research objectives presented in Chapter 3. Major findings of the research, limitations of the research, and future areas of this research for further expansion are also discussed in this chapter.

6.1 Concluding Remarks

This research proposed a new DEA-based approach for benchmarking energy efficiency in buildings in the multifamily sector and addressed the major limitations of existing DEA models. Five research objectives were presented in Chapter 3, and these research objectives and a summarized discussion for each of them are presented as follows:

Research objective 1: To find a method that remediates missing or incorrect values for instances in the dataset

A method utilizing multiple linear regression technique was proposed in this research to remediate data errors, such as missing or incorrect values. 20 DMUs with only one output error were remediated, and they can be utilized for further efficiency analysis instead of being deleted.

Research objective 2: To establish a mechanism that accurately and effectively detects outliers in the dataset

A mechanism based on the occurrence and cumulative weight of each DMU when constructing the virtually efficient DMUs was proposed in this research to detect super-efficient outliers in the dataset. Different from traditional outlier detection method, such as data cloud analysis, the proposed method considers all input and output variable at the same time, and identifies outlier based on its impact on the efficient frontier. Super-inefficient outliers are also taken into consideration in this research.

Research objective 3: To select appropriate variables to be included in the DEA model and provide justification for the selection

Although the appropriateness of variable selection is hard to be tested (Hui and Wan, 2013), three principles to of variable selection for DEA method were proposed in this research: literature review, consultation of industrial expert, and consideration of DEA assumptions. Most important, this research pointed out that EUI, a commonly used input variable in past research, is not an appropriate variable to be included in the model with corresponding justifications provided as well.

Research objective 4: To build up a DEA model that differently handles controllable variables and non-controllable variables

One of the common limitation of past research is to treat controllable and non-controllable variables of DEA model in the same way. A DEA model that handles the two types of variables different is proposed in this research. This makes very much realistic sense because non-controllable variables, such as weather impact, are simply out of property mangers' control and are therefore not a factor can be scaled up or scaled down.

Research objective 5: To quantitatively measure the stability of efficiency scores of each DMU across the entire period

Window analysis is introduced and implemented in this research to measure the stability and sensitivity of efficiency scores of DMUs across the entire period. As the results shown in Chapter 5.6, efficiency scores are subject to change when the performance of other DMUs are changing, and window analysis provides information of how stable the currently received efficiency score actually is.

The new DEA model is applied to benchmark energy efficiency in 124 buildings in the multifamily sector considering factors representing total energy consumption, building characteristics, and local weather conditions. This research contributes to the state of practice through providing a new energy benchmarking tool to facility managers and building owners that strive to relatively rank the energy-efficiency of their properties and identify low-performing properties as investment targets to enhance energy efficiency.

The entire modeling and analytical process is conducted in Excel, making it easier for facility managers and building owners to replicate the process and benchmarking properties under their management by themselves. The variables to be selected in the model can be adjusted according to different variable selection priority and data availability.

6.2 Limitations

One limitation of the proposed research is on variable selection. As mentioned in both Section 4.1 and Section 5.1, variable selection of this research is based on literature review and subject matter experts' opinions. But the final decision of variable selection is limited by the data availability. For example, the type and number of HVAC system in the building is an important factor that can influence building energy efficiency, but it was not selected in this research due to data unavailability.

Because DEA method is a peer-to-peer comparison, it does not create a fixed parametric model for benchmarking purpose, and the shape of the efficient frontier identified by the method is also subject to change as new data points come in. This feature of DEA method can create concerns as DMUs are not always compared with the same efficient frontier, but reasonable justifications can also be provided that each DMU is always compared with the top performers (efficient frontier) within its peer group. On the other hand, the regression method creates the regression model as a fixed baseline for benchmarking purpose, and building energy efficiency is evaluated based on its relative performance to that fixed baseline.

One limitation of using a fixed baseline, such as a regression model as adopted by Energy Star Portfolio Manager, for benchmarking purpose is that it is still unclear how often the fixed baseline should be updated. As Energy Star Portfolio Manager creates regression model based on survey data provided by an external organization, the update of the fixed baseline is also limited by data availability. The other limitation of using a fixed baseline is that property managers are not sure whether the appropriate selection process is taken to select a peer group of properties for benchmarking, and whether their properties are really evaluated against similar properties.

6.3 Future Research

As previously mentioned, the proposed research is limited by data availability. Therefore, one area of future work is to collect more data and try adding different relevant variables to enhance the quality of the model. Potential variables can be primarily classified into two folds: features of property and features of tenants. On the property side, variables may include property age, number of bathrooms, number of clothes dryers, number and type of HVAC system, number and type of lightening, etc., as these are the factors that would have large impact on building energy consumption. On the tenant side, variables may include occupancy rate, i.e. the presence and number of occupants, which is one of the most factors important energy efficiency of HVAC systems (Yang and Becerik-Gerber, 2016). More interestingly, it can also consider the factor of human social behavior on building energy efficiency, which represents a significant untapped potential for end-use building energy efficiency improvement (Lopes et al., 2012).

Another area of future work is to explore the reasons of significant efficiency change. As in the example of DMU 3 represented in Section 5.5, the analytical results suggest that overall efficiency score dropped and experienced significant volatilities in December 2009. It was further inferred that it might be because other DMUs had significant improvements in that month and the increase of energy consumption for DMU 3 is more significant than its peers as the weather gets cold. But the influencing factors behind the analytical results remain undiscovered and discussions and interviews with the property manager are further needed.

APPENDIX A. EFFICIENCY SCORE DISTRIBUTION BETWEEN FEBRUARY AND DECEMBER

This appendix shows results of the number of properties for three efficiency scores in different ranges for the rest of the year, namely from February 2009 to December 2009. In fact, several other results can also be presented such as efficiency score at individual property level from time-series perspective. I find this efficiency score distribution particular important because it provides the property manager an overview of all properties under management from time-series perspective.

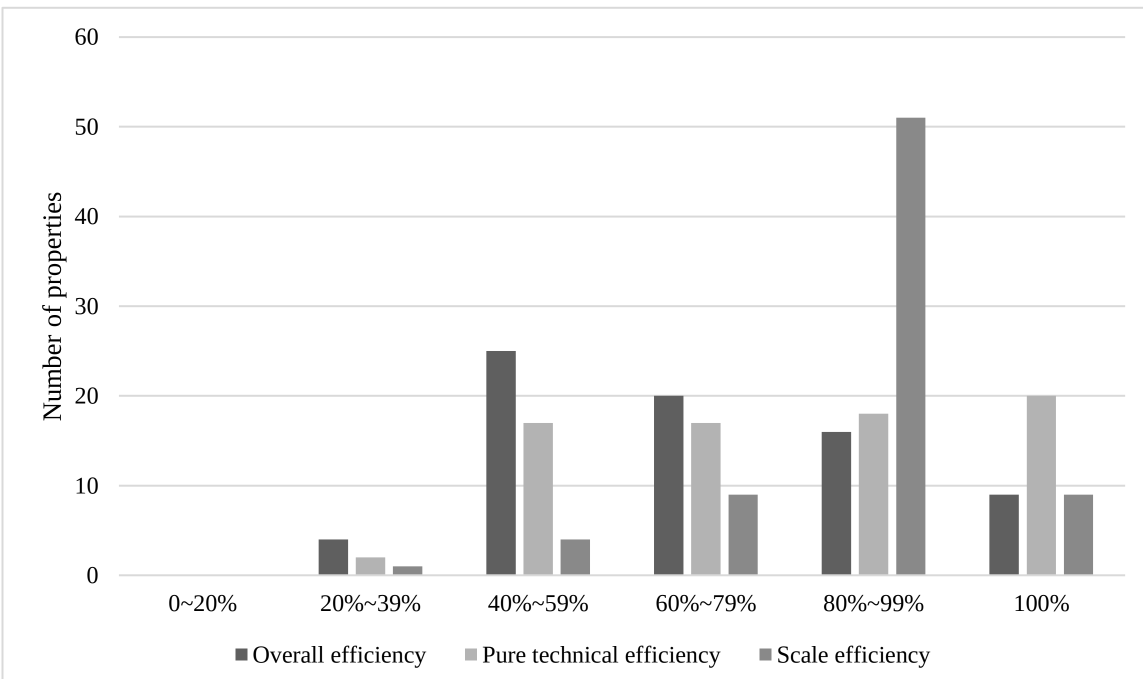

Figure A1 – Number of properties for three efficiency scores in different ranges in February 2009

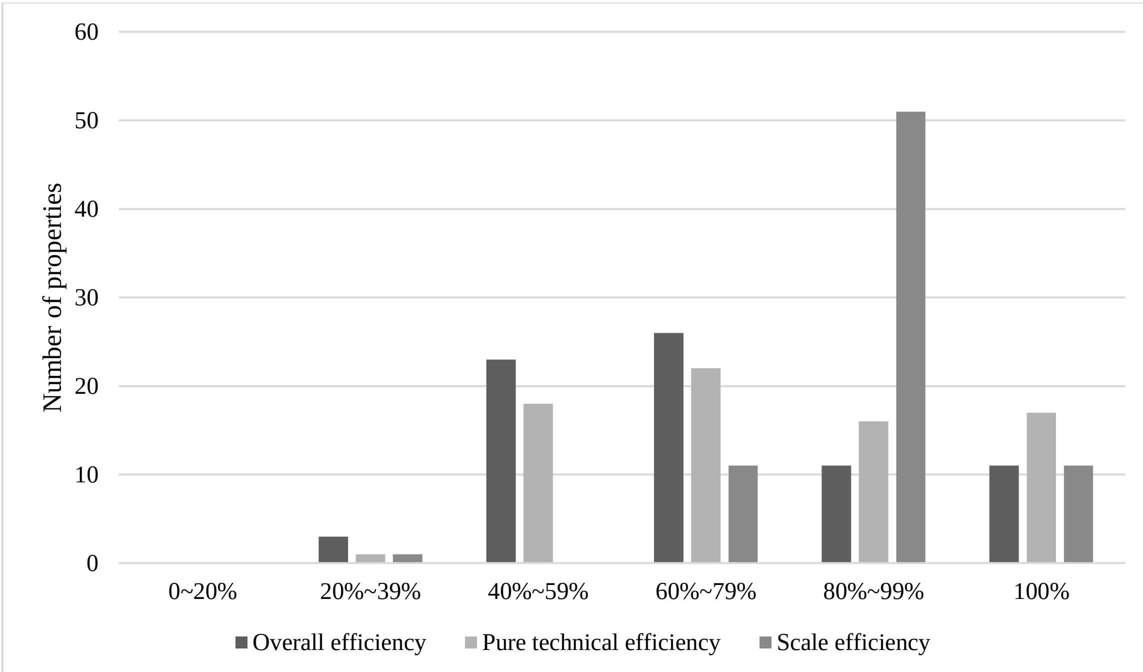

Figure A2 – Number of properties for three efficiency scores in different ranges in March 2009

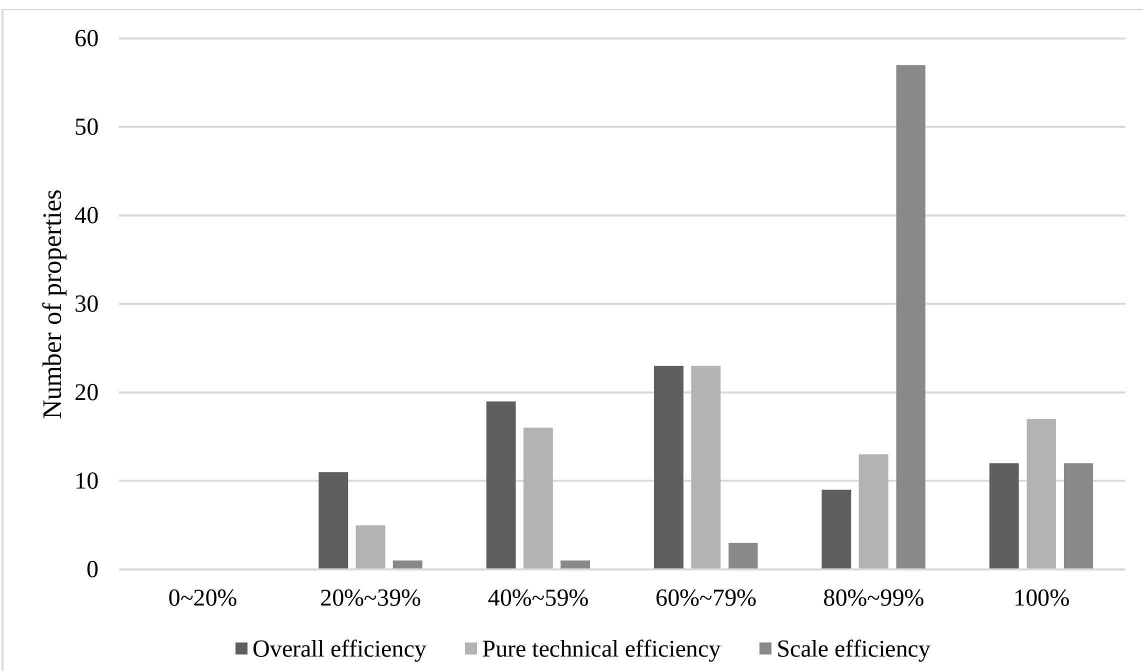

Figure A3 – Number of properties for three efficiency scores in different ranges in April 2009

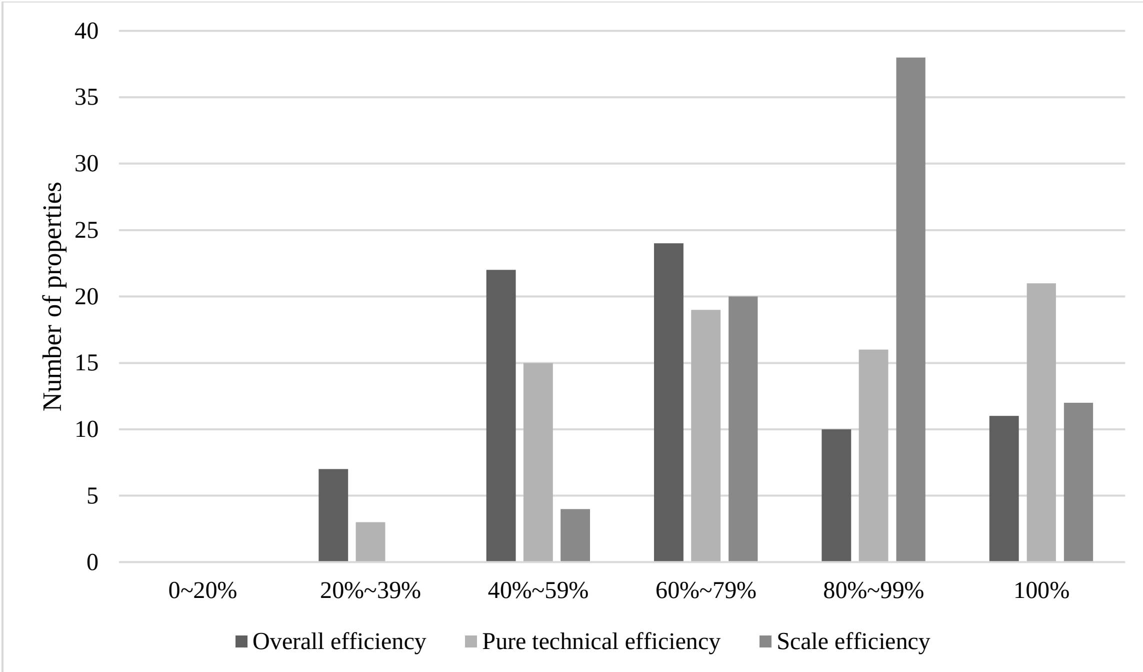

Figure A4 – Number of properties for three efficiency scores in different ranges in May 2009

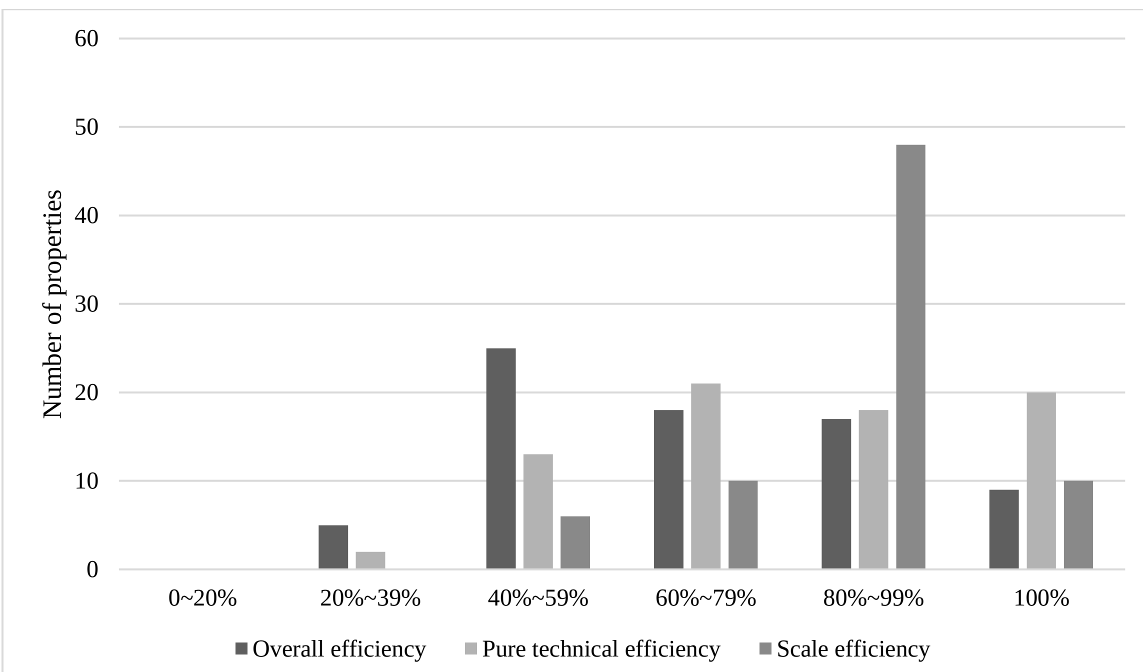

Figure A5 – Number of properties for three efficiency scores in different ranges in June 2009

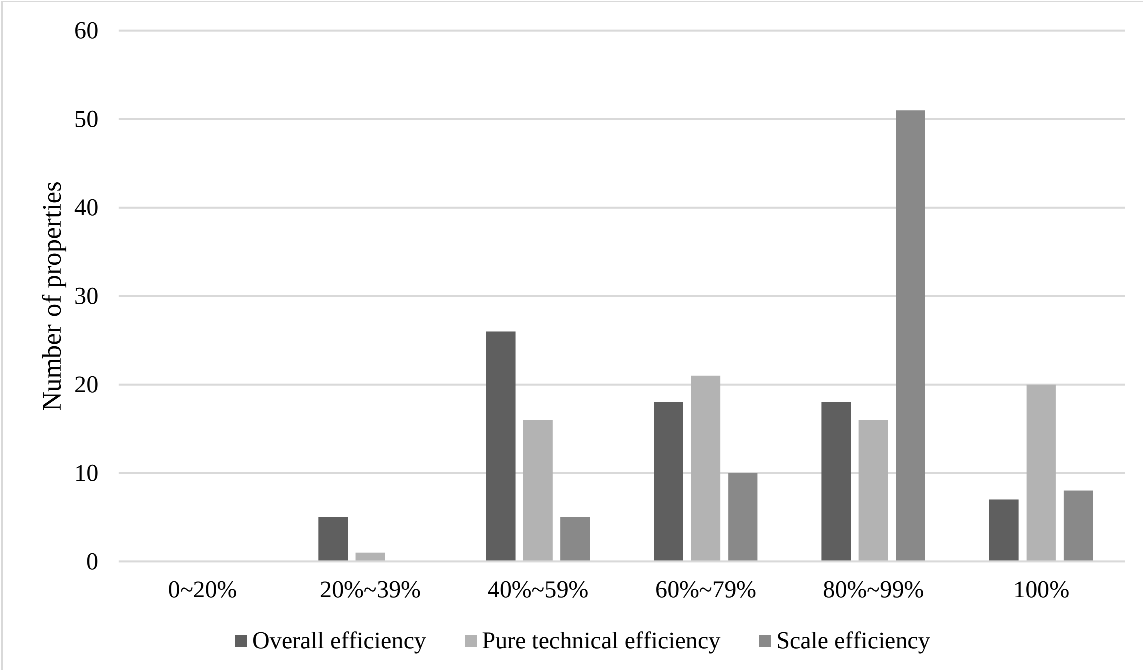

Figure A6 – Number of properties for three efficiency scores in different ranges in July 2009

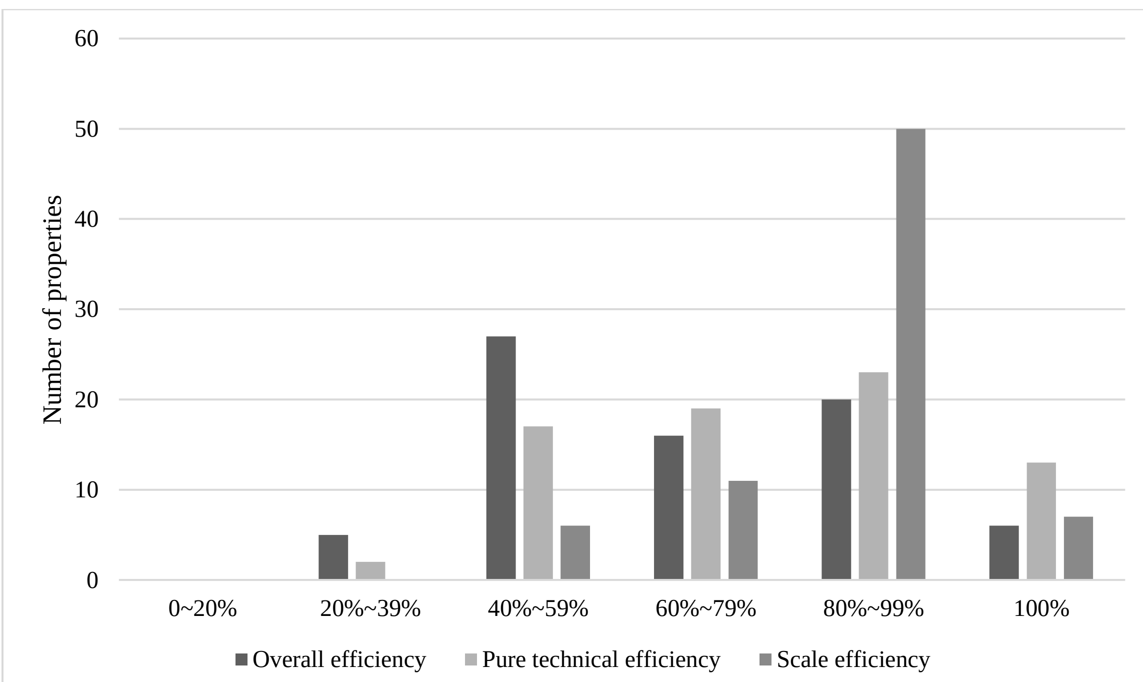

Figure A723 – Number of properties for three efficiency scores in different ranges in August 2009

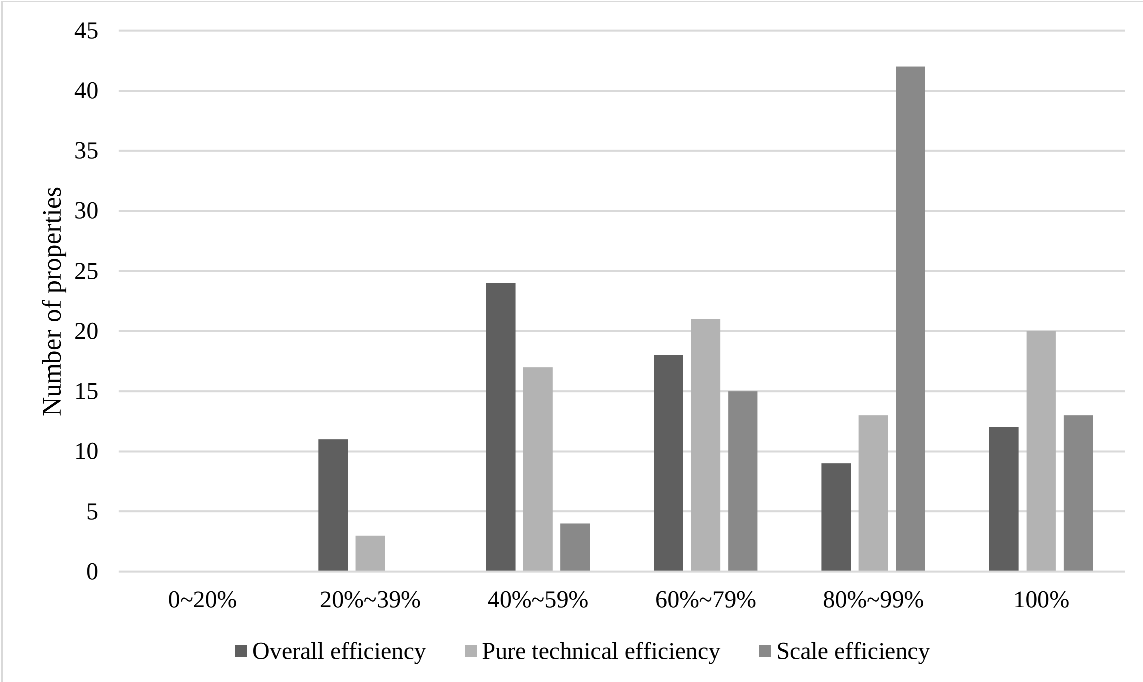

Figure A8 – Number of properties for three efficiency scores in different ranges in September 2009

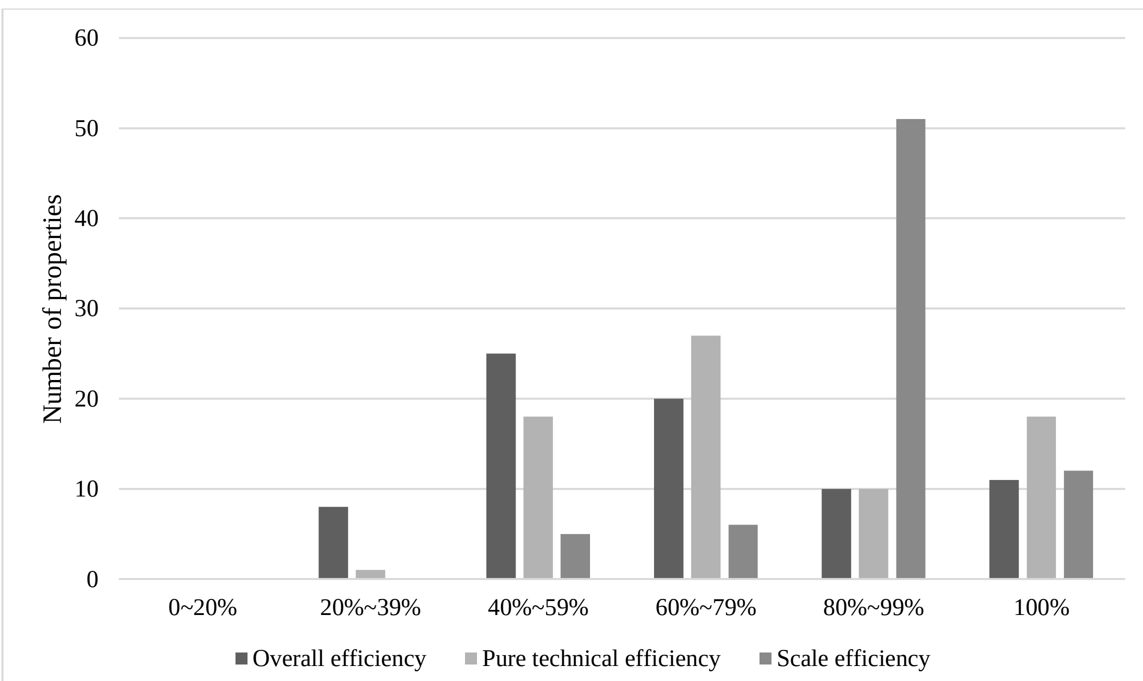

Figure A9 – Number of properties for three efficiency scores in different ranges in October 2009

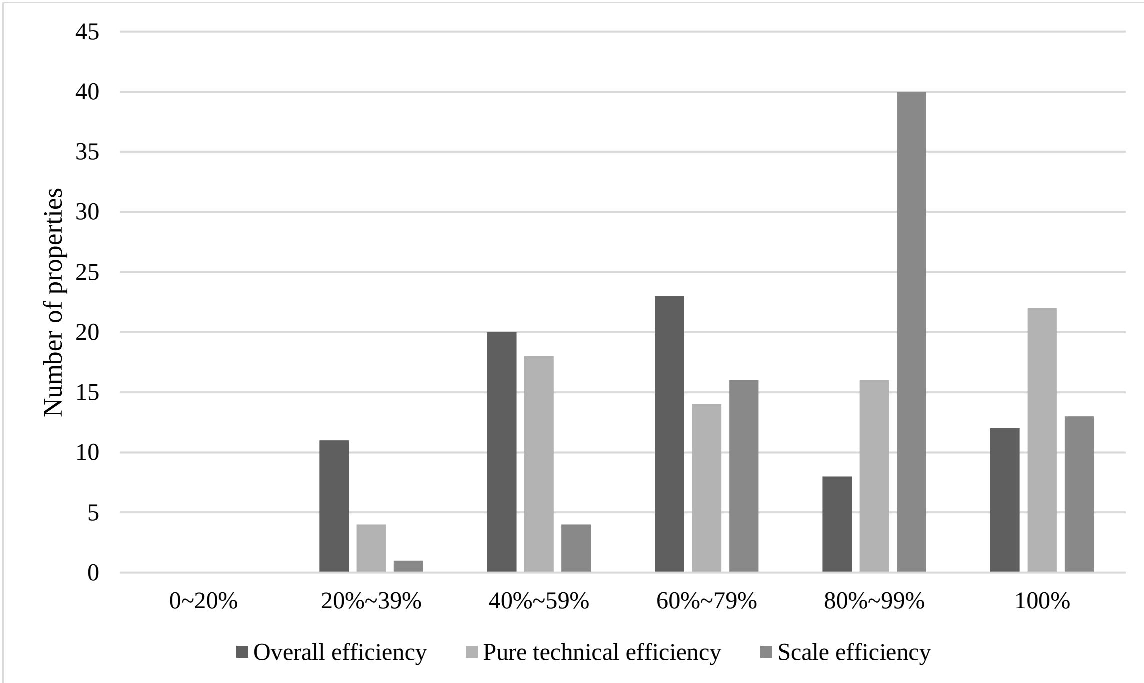

Figure A10 – Number of properties for three efficiency scores in different ranges in November 2009

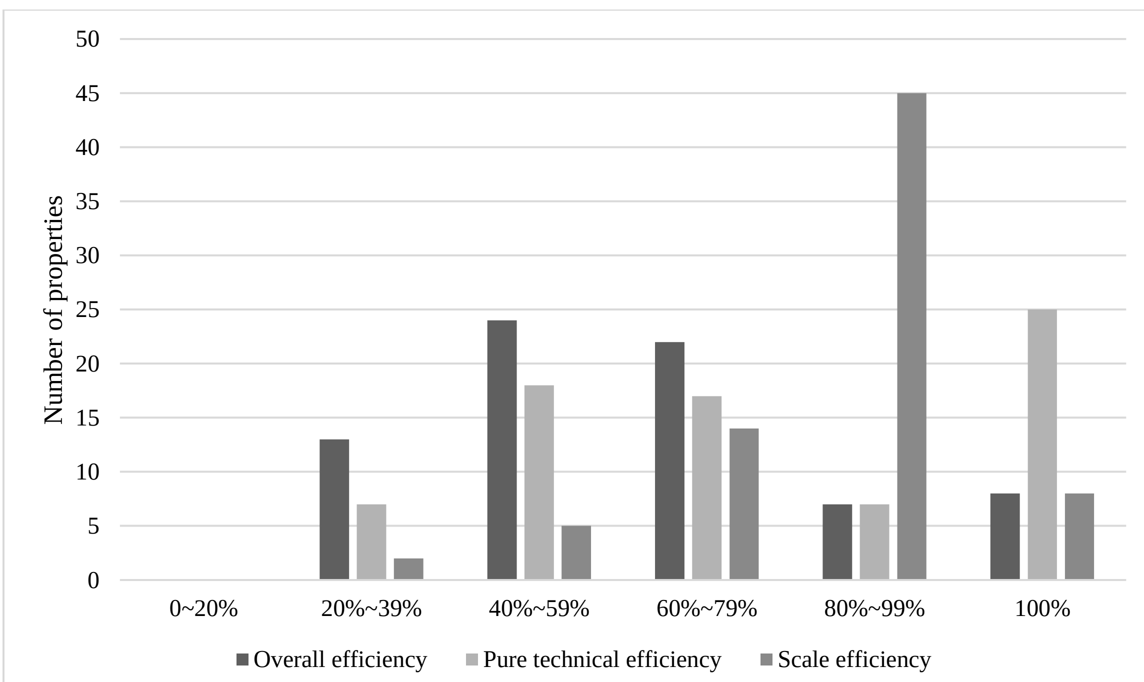

Figure A11 – Number of properties for three efficiency scores in different ranges in December 2009

REFERENCES

Amato, A.D., Ruth, M., Kirshen, P., and Horwitz, J. (2005). "Regional Energy Demand Responses to Climate Change: Methodology and Application to the Commonwealth of Massachusetts." *Climatic Change*, 71 (1-2), 175-201.

ASHRAE (2009). "Building Energy Quotient: Promoting the Value of Energy Efficiency in the Real Estate Market", < http://www.unep.org/sbci/pdfs/Paris-ASHRAE_briefing.pdf> (April, 2016)

ASHRAE (2015). "The ASHRAE Building Energy Labeling Program", < https://www.sandiego.gov/sites/default/files/legacy/environmental-services/energy/programsprojects/seab/pdf/ASHRAEpresentation.pdf> (June, 2016)

Avkiran, N. (2006). "Productivity Analysis in the Service Sector with Data Envelopment Analysis." *SSRN 2627576*, Chapter 11, 111-113.

Banker, R.D., Charnes A., and Cooper, W.W. (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science*, 30 (9), 1078-1092.

Banker, R.D., Cooper W.W., Seiford, L.M., Thrall, R.M., and Zhu, J. (2004). "Returns to Scale in Different DEA Models." *European Journal of Operational Research*, Vol. 154, 345-362.

Birtles, A.B., and Grigg, P. (1997). "Energy Efficiency of Buildings: Simple Appraisal Method." *Building Services Engineering Research and Technology*, Vol. 18, 109-114.

Bordass, W. (2005). "Onto the Radar: How Energy Performance Certification and Benchmarking Might Work for Non-Domestic Buildings in Operation, Using Actual Energy Consumption." < http://energyiq.lbl.gov/EnergyIQ/SupportPages/pdf/Onto-The-Radar-Final.pdf/> (April, 2016)

Carriere, M., Schoenau, G.J., and Besant, R.W. (1999) "Investigation of Some Large Building Energy Conservation Opportunities using the DOE-2 Model." *Energy Conservation and Management*, 40 (8), 861-872.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research*, 2(6), 429-444.

Chauhan, N.S., Mohapatra, P.K., and Pandey, K.P. (2006). "Improving Energy Productivity in Paddy Production Through Benchmarking – An Application of Data Envelopment Analysis." *Energy Conversion and Management*, 47(9), 1063-1085.

Chung, W., Hui, Y.V., and Lam, Y.M. (2006). "Benchmarking the Energy Efficiency of Commercial Buildings." *Applied Energy*, Vol. 83, 1-14.

Cooper, W.W., Seiford, L.M., and Tone, K. (2006). "Introduction to Data Envelopment Analysis and Its Uses: with DEA-Solver Software and Reference." *Springer Science & Business Media*.

Cooper, W.W., Seiford, L.M., and Zhu, J. (2011). "Handbook on Data Envelopment Analysis." *Springer Science & Business Media*.

Cullinane, K., Song, D.W., Ji, P., and Wang, T.F. (2004). "An Application of DEA Windows Analysis to Container Port Production Efficiency." Review of Network Economics, Vol. 3, Issue 2, 184-206.

Emrouznejad A. and Amin, G.R. (2009). "DEA Models for Ratio Data: Convexity Consideration." *Applied Mathematical Modeling*, Vol. 33, 486-498.

Energy Star (2008). Energy Star Building Manual, <http://www.energystar.gov/ia/business/EPA_BUM_Full.pdf> (April, 2016)

Energy Star (2014 a). "Energy Star Portfolio Manager: Technical Reference – Energy Star Score", < https://portfoliomanager.energystar.gov/pdf/reference/ENERGY%20STAR%20Score.pdf> (April, 2016)

Energy Star (2014 b). “Energy Star Portfolio Manager: Technical Reference – Energy Star Score for Hospitals in the United States”, < https://www.energystar.gov/sites/default/files/tools/Hospital_US_0.pdf> (July, 2016)

Energy Star (2014 c). “Energy Star Portfolio Manager: Technical Reference – Energy Star Score for Multifamily Housing in the United States”, < https://www.energystar.gov/sites/default/files/tools/Multifamily.pdf> (Dec, 2016)

EPBD (2003). “Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings.” <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:001:0065:0071:EN:PDF> (April, 2016)

Federspiel, C., Zhang, Q., and Arens, E. (2002) “Model-based Benchmarking with Application to Laboratory Buildings.” *Energy and Buildings*, 34 (3), 203-214.

Filippın, C. (2000). “Benchmarking the Energy Efficiency and Greenhouse Gases Emissions of School Buildings in Central Argentina.” *Building and Environment*, 35(5), 407-414.

Grösche, P. (2009). “Measuring Residential Energy Efficiency Improvements with DEA." *Journal of Productivity Analysis*, 31(2): 87-94.

Grubbs, F.E. (1969). “Procedures for Detecting Outlying Observations in Samples." *Technometrics*, 11(1): 1-21.

Hernandez, P., Burke, K., and Lewis, J.O. (2008) “Development of Energy Performance Benchmarks and Building Energy Ratings for Non-Domestic Buildings: An Example for Irish Primary Schools.” Energy and Buildings, Vol. 40 (3), 249-254.

Hui, S. C. and Wan, M. C. (2013). “Study of Hotel Energy Performance using Data Envelopment Analysis.” 12. *International Conference on Sustainable Energy Technologies*, Hong Kong, China.

Jager, J. (1983). “Climate and Energy Systems: A Review of Their Interactions.” *John Wiley & Sons*.

Khezrimotlagh, D. (2015). “How to Detect Outliers in Data Envelopment Analysis by Kourosh and Arash Method.” *ArXiv Preprint ArXiv*: 1503.02548.

Lee, W.S. (2008) “Benchmarking the Energy Efficiency of Government Buildings with Data Envelopment Analysis.” *Energy and Buildings*, 50 (5), 891-895.

Lee, W.S. and Lee, K.P. (2009) “Benchmarking the Performance of Building Energy Management using Data Envelopment Analysis.” *Applied Thermal Engineering*, Vol. 29, 3269-3273.

Lopes, MAR., Antunes, CH., and Martins, N. (2012). “Energy Behaviours as Promoters of Energy Efficiency: A 21st Century Review.” *Renewable and Sustainable Energy Reviews*, 16(6), 4095-4104.

Lu, J., Ashuri, B., and Shahandashti, M. (2014). “A Data Envelopment Analysis Model for Building Energy Efficiency Benchmarking.” *Construction Research Congress*, Atlanta, GA, 1073-1082.

Mas-Colell, A., Whinston, M.D., and Green, J.R. (1995). “Macroeconomic Theory.” *Oxford University Press*.

Olofsson, T., Meier, A., and Lamberts, R. (2004). “Rating the Energy Performance of Buildings.” The International Journal of Low Energy and Sustainable Buildings, 3.

Önüt, S. and Soner, S. (2006). “Energy Efficiency Assessment for the Antalya Region Hotels in Turkey.” *Energy and Buildings*, 38(8): 964-971.

Pacific Northwest National Laboratory (PNNL) (2012), “2011 Buildings Energy Data Book”, <http://buildingsdatabook.eren.doe.gov/docs/DataBooks/2011_BEDB.pdf/> (April, 2016).

Pérez-Lombard, L., Ortiz, J., González, R., and Maestre, I. R. (2009). “A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes.” *Energy and Buildings*, 41(3), 272-278.

Pérez-Lombard, L., Ortiz, J., and Pout, C. (2008). “A review on buildings energy consumption information.” *Energy and Buildings*, 40(3), 394-398.

Sharp, T. (1996). “Energy Benchmarking in Commercial Office Buildings.” *Proceedings of ACEEE*, Vol. 7996.

Sherman, H.D. and Gold, F. (1985). “Bank Branch Operating Efficiency Evaluation with Data Envelopment Analysis.” *Journal of Banking and Finance*, Vol. 9, 297-315.

Smirlis, Y, Maragos, E.K., and Despotis, D.K. (2006). “Data Envelopment Analysis with Missing Values: An Interval DEA Approach.” *Applied Mathematics and Computation*, Vol. 177, 1-10.

Tran, N.A., Shively, G., and Preckel, P. (2010). “A New Method for Detecting Outliers in Data Envelopment Analysis.” *Applied Economics Letters*, 17(4), 313-316.

U.S. Energy Information Administration (EIA) (2016), <http://www.eia.gov/> (April, 2016).

Wang, E., Shen, Z., Alp, N., and Barry, N. (2015). “Benchmarking Energy Performance of Residential Buildings Using Two-Stage Multifactor Data Envelopment Analysis with Degree-Day Based Simple-Normalization Approach.” *Energy Conversion and Management*, Vol. 106, 530-542.

Weather Data Depot (WDD) (2016), <http://www.weatherdatadepot.com/> (April, 2016).

Wu, X., Rajagopalan, P., and Lee, S.E. (2010). “Benchmarking Energy Use and Greenhouse Gas Emissions in Singapore’s Hotel Industry.” *Energy Policy*, Vol. 38, 4520-4527.

Yang, Z. and Becerik-Gerber, B. (2016). “How Does Building Occupancy Influence Energy Efficiency of HVAC Systems” *Energy Procedia*, Vol. 88, 775-780.

Yue, P. (1992). “Data Envelopment Analysis and Commercial Bank Performance: A Primer with Applications to Missouri Banks” *Review*, Vol. 74, 31-45.

VITA

Jun Wang

Jun Wang was born in Chongqing, China. He received his Bachelor’s degree in Construction Management from Tianjin University, China in 2012. He attended the Georgia Institute of Technology, Atlanta, U.S.A. from Aug. 2012 to May 2017 to pursue his doctorate in Civil Engineering and two Master of Science degrees with one in Quantitative and Computational Finance and the other one in Computer Science. Jun’s research areas can be summarized into three categories: construction safety, construction cost index forecasting, and building energy benchmarking. By the time of his graduation, Jun’s research findings will have been presented in multiple journal publications and conference proceedings.