FRAMEWORK TO EVALUATE ENTROPY BASED DATA FUSION METHODS IN

SUPPLY CHAIN MANAGEMENT

Huong Thi Tran, B.S., Ph.D.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2016

APPROVED:

Robert Pavur, Committee Chair
David Nowicki, Committee Co-Chair
Terrance Pohlen, Committee Member
Jeff Sager, Chair of the Department of
Marketing and Logistics
Marilyn Wiley, Dean of the College of
Business
Victor Prybutok, Vice Provost of the
Toulouse Graduate School

Tran, Huong Thi. *Framework to Evaluate Entropy Based Data Fusion Methods in Supply Chain Management*. Doctor of Philosophy (Business Administration-Logistics Systems), December 2016, 96 pp., 31 tables, 3 figures, references, 145 titles.

This dissertation explores data fusion methodology to deduce an overall inference from the data gathered from multiple heterogeneous sources. Typically, if there existed a data source in which the data were reliable and unbiased, then data fusion would not be necessary. Data fusion methodology combines data form multiple diverse sources so that the desired information - such as the population mean - is improved despite redundancies, inaccuracies, biases, and inflated variability in the data. Examples of data fusion include estimating average demand from similar sources, and integrating fatality counts from different media sources after a catastrophe.

The approach in this study combines "inputs" from distinct sources so that the information is "fused." Another way of describing this process is "data integration." Important assumptions are 1. Several sources provide "inputs" for information used to estimate parameters of a probability distribution. 2. Since distributions for the data from the sources are heterogeneous, some sources are less reliable. 3. Distortions, bias, censorship, and systematic errors may be more prominent in data from certain sources. 4. The sample size of sources data, number of "inputs," may be very small.

Examples of information from multiple sources are abundant: traffic information from sensors at intersections, multiple economic indicators from various sources, demand data for product using similar retail stores as sources, polling data from various sources, and disaster count of fatalities from different media sources after a catastrophic event.

This dissertation seeks to address a gap in the operations literature by addressing three research questions regarding entropy base data fusion (EBDF) approaches to estimation. Three separate, but unifying, essays address the research questions for this dissertation.

Essay 1 provides an overview of supporting literature for the research questions. A numerical analysis of airline maximum wait time data illustrates the underlying issues involved in EBDF methods. This essay addresses the research question: Why consider alternative entropy-based weighting methods?

Essay 2 introduces 13 data fusion methods. A Monte Carlo simulation study examines the performance of these methods in estimating the mean parameter of a population with either a normal or lognormal distribution. This essay addresses the following research questions: 1. Can an alternative formulation for Shannon's entropy enhance the performance of Sheu (2010)'s data fusion approach? 2. Do symmetric and skewed distributions affect the 13 data fusion methods differently? 3. Do negative and positive biases affect the performance of the 13 methods differently? 4. Do entropy based data fusion methods outperform non-entropy based data fusion methods? 5. Which data fusion methods are recommended for symmetric and skewed data sets when no bias is present? What is the recommendation under conditions of few data sources?

Essay 3 explores the use of the data fusion method estimates of the population mean in a newsvendor problem. A Monte Carlo simulation study investigates the accuracy of the using the estimates provided in Essay 2 as the parameter estimate for the distribution of demand that follows an exponential distribution. This essay addresses the following research questions: 1. Do data fusion methods with relatively strong performance in estimating the parameter mean estimate also provide relatively strong performance in estimating the optimal demand under a given ratio of overage and underage costs? 2. Do any of the data fusion methods deteriorate or improve with the introduction of positive and negative bias? 3. Do the alternative entropy formulations to Shannon's entropy enhance the performance of the methods on a relative basis? 4. Is the relative rank ordering performance of the data fusion methods different in Essay 2 and Essay 3 in the resulting performances of the methods?

The contribution of this research is to introduce alternative EBDF methods, and to establish a framework for using EBDF methods in supply chain decision making. A comparative Monte Carlo simulation analysis study will provide a basis to investigate the robustness of the proposed data fusion methods for estimation of population parameters in a newsvendor problem with known distribution, but unknown parameter. A sensitivity analysis is conducted to determine the effect of multiple sources, sample size, and distributions.

# ACKNOWLEDGMENTS

I would like to thank God for being my source of strength, my family for their love, and my committee, Dr. Robert Pavur, Dr. David Nowicki, and Dr. Terrance Pohlen, for their valuable insights.

Specifically, I would like to thank my advisor, Dr. Robert Pavur, without whom this dissertation would not have been completed, for his tremendous expertise, patience, and support.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ..... iii

LIST OF TABLES ..... vii

LIST OF FIGURES ..... ix

CHAPTER 1 INTRODUCTION ..... 1

1.1. Importance of Data Fusion Methods ..... 1

1.2. Examples of Applications of Data Fusion Supporting Relevance of Research ..... 5

1.2.1. Data Fusion in Defense and Aerospace Industries ..... 5

1.2.2. Data Fusion in Robotics and Machine Intelligence ..... 6

1.2.3. Data Fusion in Remote Sensing ..... 6

1.2.4. Data Fusion in Vehicular Traffic Management ..... 7

1.2.5. Data Fusion in SCM Decision Theoretic Models ..... 7

1.2.6. Data Fusion in Logistical Post-Disaster Relief Efforts ..... 9

1.3. Identification of the Research Problem ..... 9

1.4. Statement of Purpose ..... 12

1.5. Essays Comprising Research ..... 13

1.5.1. Essay 1 Research Question ..... 14

1.5.2. Essay 2 Research Questions ..... 15

1.5.3. Essay 3 Research Questions ..... 16

1.6. Methodology to Investigate Research Questions ..... 17

1.7. Theoretical and Practical Contributions, and Summary of Conclusions from Research ..... 19

1.7.1. Theoretical Contributions ..... 19

1.7.2. Practical Contributions ..... 20

1.7.3. Essay 1 Summary of Conclusions ..... 20

1.7.4. Essay 2 Summary of Conclusions ..... 21

1.7.5. Essay 3 Summary of Conclusions ..... 22

1.8. Overview of Dissertation Organization ..... 23

CHAPTER 2 ESSAY 1: NUMERICAL COMPARISON OF DATA FUSION METHODS USING TWO ENTROPY BASED METHODS ..... 25

2.1. Abstract ........ 25
2.2. Introduction to Entropy-based Methodology ........ 25
2.3. The Role of Data Fusion in SCM Applications ........ 27
2.3.1. Assessing System Risk ........ 28
2.3.2. Forecasting Measures of Logistics ........ 29
2.3.3. Enhancing Systems ........ 30
2.3.4. Improving the Newsvendor Model's Inventory Estimation ........ 30
2.3.5. Summary of Entropy Application for SCM Models ........ 31
2.4. Importance of Entropy-Based Data-Fusion (EBDF) Framework in SCM ........ 32
2.4.1. Data Issues in SCM ........ 33
2.4.2. Multiple Sources in SCM Operations ........ 33
2.4.3. Data Bias and Distortion ........ 34
2.4.4. Censored Data ........ 34
2.5. Probabilistic Formulation of Entropy Based Data Fusion ........ 35
2.6. Intervals for Belief-Strength Bands Used to Provide Posterior Probabilities ........ 38
2.7. Numerical Illustration of Decision Making with Entropy Based Data Fusion ........ 40
2.8. Contrasting Shannon's and Pal & Pal's Entropy Formulations ........ 44
2.9. Conclusion ........ 46

CHAPTER 3 ESSAY 2: PARAMETER ESTIMATION: COMPARATIVE PERFORMANCE OF PROPOSED ENTROPY AND NON-ENTROPY BASED DATA FUSION ESTIMATORS ........ 47

3.1. Abstract ........ 47
3.2. Introduction ........ 47
3.3. Proposed Entropy Based Data Fusion Methods ........ 48
3.4. Simulation Study Procedure to Assess Data Fusion Methods ........ 51
3.5. Monte Carlo Simulation Results of Data Fusion Estimation of Mean Using Normally Distributed Inputs with Possible Bias ........ 53
3.6. Simulation Results of Data Fusion Estimation of Mean Parameter Using Lognormally Distributed Inputs with Possible Bias ........ 59
3.7. Conclusions for Data Fusion Methods Estimating the Population Mean ........ 65

CHAPTER 4 ESSAY 3: NEWSVENDOR MODEL: COMPARATIVE PERFORMANCE OF PROPOSED ENTROPY AND NON-ENTROPY BASED DATA FUSION ESTIMATORS ........ 67

4.1. Abstract ........ 67

4.2. Limited Knowledge of Demand Distribution in Newsvendor Applications ........ 68
4.3. Description of Results of Newsvendor Model Performance for Methods............ 69
4.4. Newsvendor Results from Simulation Study: Data Fusion Using Normally Distributed Inputs with Possible Bias .................................................................. 70
4.5. Newsvendor Results from Simulation Study: Data Fusion Using Lognormally Distributed Inputs with Possible Bias .................................................................. 77
4.6. Conclusions for Data Fusion Methods in the Newsvendor Model ....................... 82
4.7. Limitations to Results and Conclusions................................................................ 84

REFERENCES .......................................................................................................................... 86

LIST OF TABLES

Table 1.1 Definitions of data fusion techniques in the academic literature. .................................. 4

Table 1.3 Specific SCM applications benefitting from entropy methods. .................................. 10

Table 2.1 Summary contributions of entropy based methods in SCM ....................................... 31

Table 2.4 Comparison of Shannon's and Pal & Pal's entropies for discrete uniform ................. 37

Table 2.6.1 Max wait time in minutes at Atlanta Airport - May 31, 2016 to June 3, 2016.......... 42

Table 2.6.2 Estimates of the average maximum wait time ......................................................... 43

Table 2.6.3 Maximum wait time in minutes at Atlanta Airport from May 31, 2016 to June 3, 2016.......................................................................................................................................... 43

Table 2.6.4 Estimates of the average maximum wait time under data distortion. ........................ 44

Table 2.7.1 Probability Distributions for 5 Groups .................................................................... 45

Table 2.7.2 Contrasting Sheu's Weights with Shannon's and Pal & Pal's Entropies.................. 45

Table 3.2.1 Proposed EBDF Methods Using Sheu (2010)'s Methodology................................. 49

Table 3.2.2 Data Fusion Methods Not Using Sheu (2010)'s Methodology ................................ 50

Table 3.3.1 Data are simulated from sources having either normal or lognormal distributions, with increasing standard deviations to indicate a reduction in reliability. .................................... 51

Table 3.3.2 Biases added to source 1 in simulation study. The zero positive and zero negative under each distribution is only one configuration. Bias is added only to source 1....................... 52

Table 3.3.3 Simulation study configuration conditions .............................................................. 53

Table 3.4.1 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 4 sources with 5 inputs per source.  Data follow a normal distribution. ........... 55

Table 3.4.2 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 4 sources with 15 inputs per source.  Data follow a normal distribution. .......... 56

Table 3.4.3 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 8 sources with 5 inputs per source. Data follow a normal distribution. ............ 57

Table 3.4.4 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 8 sources with 15 inputs per source. Data follow a normal distribution. .......... 58

Table 3.5.1 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 4 sources with 5 inputs from a lognormal distribution ................................................. 60

Table 3.5.2 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 4 sources with 5 inputs from a lognormal distribution ................................................ 61

Table 3.5.3 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 8 sources with 5 inputs from a lognormal distribution ................................................ 62

Table 3.5.4 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 8 sources with 15 inputs from a lognormal distribution .............................................. 63

Table 4.3.1 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 5 inputs from a normal distribution and assuming exponentially distributed demand. .................. 72

Table 4.3.2 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 15 inputs from a normal distribution and assuming exponentially distributed demand. ............. 73

Table 4.3.3 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 5 inputs from a normal distribution and assuming exponentially distributed demand. .................. 74

Table 4.3.4 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 15 inputs from a normal distribution and assuming exponentially distributed demand. ............. 75

Table 4.4.1 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 5 inputs from a lognormal distribution and assuming exponentially distributed demand. ............. 78

Table 4.4.2 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 15 inputs from a lognormal distribution and assuming exponentially distributed demand. ......... 79

Table 4.4.3 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 5 inputs from a lognormal distribution and assuming exponentially distributed demand. ............. 80

Table 4.4.4 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 15 inputs from a lognormal distribution and assuming exponentially distributed demand. ......... 81

LIST OF FIGURES

Figure 1. Entropy based models can be considered to be a subset of probability based models.. 12

Figure 2. Layout of dissertation essays. .......................................................................... 13

Figure 3. Distribution of levels of belief strength for a Gaussian population according to the empirical rule. ........................................................................................................ 39

# CHAPTER 1

# INTRODUCTION

This dissertation investigates the integration of multiple-source data information available through a stochastic process and explores data fusion methodology to deduce an overall inference from the data. A simple example of data "fusing" is combining estimates of traffic flow from two traffic sensors whose target traffic may be viewed from opposite angles. Data fusion supports supply chain managers with decision making since they often must use all information from heterogeneous sources to achieve enhanced inference about their operations.

## 1.1. Importance of Data Fusion Methods

Before the research statement and questions are presented, a description of the importance of data fusion methods to this research is discussed. This section provides the necessary background to provide a basis for further research investigation into introducing and assessing data fusion methods.

First, the term "data fusion" is a general term that is typically used in the literature to indicate that data points from several sources are "fused" to obtain improved, more relevant information. Another way of describing this process is "data integration." Important assumptions about the data collection process are:

1. Several (multi) sources provide "inputs" for information about descriptive parameters of data.

2. Heterogeneous distributions for the data from the sources make some sources less reliable.

3. Distortions, bias, censorship, outlier observations, and systemic errors may be more prominent in data from certain sources.

4. Sample size of data, the number of “inputs” from sources is generally small.

Examples of information from multiple sources abound: traffic information from sensors at intersections, multiple economic indicators from various sources, demand data for a product using similar retail stores as sources, polling data from various sources, and disaster fatality count from different media sources after a catastrophic event.

The purpose of data fusion methodology is to combine data from multiple sources in such a way that relevant information from a limited number of sources is improved despite redundancies, inaccuracies, biases, partial information, and inflated variability in the data. If a “benchmark” data source in which reliable, representative, and unbiased data existed, then data fusion would not be necessary. Ultimately, supply chain managers must often merge or integrate information from diverse sources to provide measures with a higher level of reliability. Data fusion origins can be traced to intrinsic behaviors performed by humans and animals to improve their ability to survive. This concept is analogous to the instincts of humans and animals to simultaneously utilize multiple senses for threat assessment or food collection. For example, a combination of smell, touch, sight, and taste allows an animal to examine the edibility of a new substance in a far superior manner than taste alone. Similarly, danger is more accurately detected with hearing and experience rather than with hearing alone (Hall and Llinas, 2001).

Fusion of data from multiple sensors is used to improve the observation process. For example, in radar tracking, a moving aircraft can be more accurately observed by both a pulsed radar and an infrared imaging sensor. Since the radar can only detect the aircraft’s range, but not the angular direction, and the infrared imaging sensor can provide the angular direction, but not the range, the data fusion from these two sources creates a more effective radar tracking.

Data fusion has widespread applications. Historically, the integration of data sources greatly benefited military applications such as target tracking, submarine warfare, and battlefield surveillance (Durrant-Whyte and Henderson, 2008; Hall and Llinas, 2001; Hall and McMullen, 2004). However, the last few decades have seen a rise in nondefense data fusion applications in forecasts of weather conditions, economic changes, and geopolitical activity; maintenance of complex machinery; robotics, and medical applications (Durrant-Whyte and Henderson, 2008; Hall and McMullen, 2004); and real-time traffic applications (El Faouzi et al., 2011). Applications of data fusion in logistics and supply chain management (SCM) are also on the rise. They include monitoring of manufacturing process (Hall and McMullen, 2004), assessing and predicting the performance degradation of a process (Djurdjanovic et al., 2003; Lee et al., 2006); facilitating system integration between manufacturing and office planning (Qiu, 2002); facilitating the coordination of SCM (Hsu and Wallace, 2007); and providing dynamic demand forecasting in emergency logistics (Sheu, 2010).

Many attempts to define data fusion can be found in the literature. Initially, White Jr. (1987) of the Joint Directors of Laboratories (JDL) Data Fusion Subgroup defined data fusion as “a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates.” Klein (1993) broadened this definition, stating that data can be provided by a single source with many instances or by multiple sources. In 1999, the JDL redefined data fusion as “the process of combining data to refine state estimates and predictions” in recognition of its diverse applications (Steinberg et al., 1999). A selection of important definitions of data fusion in the past few decades in both academic and industrial areas are summarized in Table 1.1. Additional literature reviews are provided in Essay 1 which discuss the importance of entropy based applications in SCM.

Table 1.1 Definitions of data fusion techniques in the academic literature.

*Summary of Definitions of Data Fusion*

| Defined by | Data Fusion Definition |
|---|---|
| White Jr., 1987 | "A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance." |
| Klein, 1993 | "A multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources." |
| Mangolini, 1994 | "A set of methods, tools and means using data coming from various sources of different nature, in order to increase the quality (in a broad sense) of the requested information." |
| Li et al., 1995 | "The combination of a group of sensors with the objective of producing a single signal of greater quality and reliability" |
| Hall and Llinas, 1997 | "Techniques that combine data from multiple sensors, and related information from associated databases, to achieve improved accuracy and more specific inferences that could be achieved by the use of a single sensor alone" |
| Pohl and Van Genderen, 1998 | "The combination of two or more different images to form a new image by using a certain algorithm" |
| Steinberg et al., 1999 | "The process of combining data to refine state estimates and predictions" |
| Wald, 1999 | "A formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application." |

Several realizations emerge from these definitions. First of all, since multisensory data fusion is a process that combines data from multiple sources, performing data fusion requires that data be both available and authentic (Khaleghi et al., 2013). Secondly, since data comes from multiple sources, data fusion is often faced with challenges such as heterogeneous data distribution and conflicting data information (Dong and Naumann, 2009; Khaleghi et al., 2010; Smith and Singh, 2006).

## 1.2. Examples of Applications of Data Fusion Supporting Relevance of Research

Initially appearing in the 1960s as mathematical models for data integration, data fusion gained importance during the 1970s in the literature illustrating its applications. The importance of data fusion increased with the advent of multi-sensors, specifically developed to capture and fuse data to enhance information quality. Currently data fusion is the focus of many research areas. This section covers the applications of data fusion in industries such as defense and aerospace, robotics and machine intelligence, remote sensing, traffic management, and SCM. The reason that several sources of data are used in the data collection process is because reliable data is often difficult to obtain. Some sources will be more reliable than others. Knowing this a priori is not always possible.

### 1.2.1. Data Fusion in Defense and Aerospace Industries

The defense industry has a natural interest in data integration since it collects data on numerous targets. In defense and aerospace applications, data fusion is a vital technique in automated target recognition, battlefield surveillance, guidance and control of autonomous vehicles, and submarine warfare. Alessandretti et al. (2007) and Blair et al. (1991) described a detection system that calibrates and then fuses together asynchronous data received by dissimilar sensors to enhance detection. Hess and Fila (2002) proposed an autonomic system in which multi-sensors capture, fuse together and then inform the supply chain the conditions of multi components of an aircraft so that necessary logistics action can be taken. Data fusion was also a technology vital to threat assessment, which is a combined evaluation of enemy intent, enemy capability, and consequences of threat activities to determine a course of action to appropriately deploy resources to neutralize or eliminate the threat. Diamond and Ceruti (2007) presented a model that fuses data collected from an aerial sensor delivery system to detect targeting discrepancies and asymmetric threats in opaque environments.

### 1.2.2. Data Fusion in Robotics and Machine Intelligence

Data fusion is used extensively in robotics and machine intelligence. Equipped with various sensors such as sonar, cameras and other technologies, a collaborative robot deployed on a manufacturing floor is capable to perform manufacturing tasks while working alongside humans without causing them injury (Fong et al., 2001; Girod and Estrin, 2001; Stroupe et al., 2001). Researchers also focused on using data fusion based systems for humanitarian de-mining, since data from one sensor is generally insufficient and ineffective for the landmine detection that meets the requirements. The terminology "data fusion" is more common in the engineering and operations disciplines than in the general business disciplines.

### 1.2.3. Data Fusion in Remote Sensing

Multi-source data fusion techniques are emerging in remote sensing to help scientists acquire information without making physical contact with the object under surveillance; for example, the scanning of the earth by satellite (Benz et al., 2004; Ehlers, 1991; Pohl and Van

Genderen, 1998; Solberg et al., 1994). The "multi" concept for remote sensing applications often refers to multisource. Data fusion research in remote sensing includes the fusion of multiple classifiers for landslide monitoring (Metternicht et al., 2005; Nichol and Wong, 2005), for classification of complex forest area (Dalponte et al., 2008), and for classification of urban areas (Fauvel et al., 2006).

### 1.2.4. Data Fusion in Vehicular Traffic Management

Data fusion models are important in vehicular traffic management. These models provide real-time traffic information (Bachmann et al., 2013; Dell'Orco andTeodorovic, 2009; El Faouzi et al., 2011; Kong, Li et al., 2009) to control traffic by estimating travel time (Cheu et al., 2001; Nelson and Palacharla, 1993; Tarko and Rouphail, 1993) ), estimating road conditions (Byon et al., 2010) , detecting incidents (Bhandari et al., 1995; Ivan, 1996; Ivan et al., 1995), or classifying different traffic states (Kong et al., 2009; Treiber et al, 2011).

### 1.2.5. Data Fusion in SCM Decision Theoretic Models

Research has demonstrated the importance of data accuracy and availability in SCM (Davis, 1993; Lee and Billington, 1995; Lummus and Vokurka, 1999). Protecting the enterprise data from malicious attacks equals protecting the competitive advantage of the supply chain. One such attacks is the denial-of-service (DoS) by perpetrators seeking to paralyze the enterprise networks and make data unavailable to users. Bass (1999, 2000) suggested creating cyberspace intrusion detection systems that fuse data from heterogeneous sensors to counter the denial-of-service attacks. Siaterlis and Maglaris (2004) and Siraj et al. (2004) presented DoS detection engines based on a data-fusion paradigm to appraise the trustworthiness of information requests.

One key aspect of SCM is the distribution logistics that includes inventory control and transportation management (Cooper and Ellram, 1993; Cooper et al., 1997; Thomas and Griffin, 1996). An ineffective inventory management system results in surpluses or shortage of bulk materials purchased, and reduces profit (Bell and Stukhart, 1986). For example, in construction supply chains, poorly identifying, tracking, and locating highly customized prefabricated components results in late deliveries, incorrect installations, schedule delays, and higher labor costs (Song et al., 2006). Razavi and Haas (2010) proposed an automated data fusion method, focusing on the detection of dislocation and multi-handling of materials to improve the precision of locating inventory. Carthel et al. (2007), Easley (2005), and Tan et al. (2010) invented data-fusion inventory controlling systems that integrate data from multi container tracking sources.

Transportation management is often difficult since it involves coordinating heterogeneous transporters and multiple demand centers without the benefit of real-time traveling times and homogeneous demand and capacity in demand centers (Lei et al., 2006). Hsu and Wallace (2007) proposed adding a digital layer, connected to the enterprise information systems and to the traveling routes, to report real-time freight information and dynamically control the routing and scheduling of the inventory.

Besides logistics, demand forecasting is another key aspect in Supply Chain Management. Accurate forecasting helps prevent inventory shortage or surplus. Pyle (2003) and Southworth et al. (2008) proposed various models to discover important features affecting demands then integrating results from these models to forecast demand. Acknowledging the growing importance of data fusion concepts, Khan et al. (2008) suggested decision models that combine data mining with data fusion to transform raw data to forecasting insights.

### 1.2.6. Data Fusion in Logistical Post-Disaster Relief Efforts

Data fusion has also been a focus of research in humanitarian and relief logistics. Llinas (2002) described the overall strategic approach addressing information fusion to support crisis-center decision makers in post-disaster environments for both natural and man-made disasters. Scott and Rogova (2004) constructed a post-crisis management scenario to explore various data fusion system designs that could filter data from redundant or contradictory reports. Gong et al. (2004) offered a post-disaster scalable decision-making methodology that requires support of data fusion. Jotshi et al. (2009) developed a methodology that integrates data for routing and dispatching emergency vehicle in the aftermath of an earthquake. Sheu (2010) proposed a dynamic relief-demand management model for emergency logistics operations under imperfect information conditions to forecast relief demand.

## 1.3. Identification of the Research Problem

This research is inspired by SCM applications that require robust methods to make sense of information collected through multi-sources usually driven by technology. How do supply chain managers combine, integrate, and fuse data from diverse sources in an age of time-sensitive data available through technology-driven sources, sometimes with unknown reliability? This research investigates robust methods of weighting sources of information that may be critical to SCM decisions. Entropy based methodology allows decisions to be weighted by a measure of uncertainty. Entropy approaches in data fusion methods have not received sufficient attention in the SCM literature and research is needed to provide a framework for its use.

Since this dissertation’s problem statement involves researching entropy based methods, then there needs to be a derived SCM benefit. What are the decision making applications in SCM that entropy based methods can enhance?  Well known applications include the newsvendor model, inventory management, and transportation strategies. Table 1.3 names specific decision making applications in SCM and list research articles that investigate entropy based methods to enhance the application. As noted in Table 1.3, additional SCM applications benefiting from entropy based methods include: forecasting demand, personnel staffing, assessing supplier risks, simulating and modeling emergency relief efforts, and optimizing business processes and online logistical support. Limited information and knowledge of uncertainty makes entropy based approaches appealing.

Table 1.3 Specific SCM applications benefitting from entropy methods.

| SCM Applications Benefiting from Entropy Based Methodology | Supplier Risk & Procurement | Optimizing Business Process | Forecasting & Predicting Demand | Staff Planning | Transportation Strategies | Online Logistics & E-Commerce | Product & Service Operations | Newsvendor | Inventory Management | SCM Modeling & Simulation |
|---|---|---|---|---|---|---|---|---|---|---|
| Allesina et al. (2010) | X | X | | X | | | | | | |
| Andersson et al. (2013) | | | | | | | | X | X | X |
| Arkhipov and Ivanov (2011) | X | X | | | | | | | | |
| Chen and Freeman (2014) | X | X | X | | | | | | | X |
| Cheng et al. (2006) | | | X | | X | | | | | X |
| Dekkers et al. (2012) | | X | | | | | X | | X | |
| Eren and Maglaras (2006) | | | X | | | | X | | | |
| Gan and Wirth (2005) | | | | X | | X | | | | X |
| Ghorbani et al. (2012) | X | | | X | | | X | | | |
| Guoyi and Xiaohua (2011) | X | | | | X | | | | X | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hu et al. (2008) | X | | | | | | X | | | X |
| Sheu (2010) | | | X | | | | X | | | |
| Zheng et al. (2013) | | | | | | X | X | | | |

As Sheu (2010)'s study illustrates, a timely information-based decision is required for expedient responses to urgent relief challenges. His paper presented a hybrid fuzzy clustering-estimation approach. Andersson et al. (2013) considered the newsvendor problem under partial information. One approach to the newsvendor problem is to use conservative rules and assume full information about the distribution of the demand. In this case, solutions are straight-forward. Although the newsvendor problem has been studied under partial information with a variety of extensions, a data fusion entropy approach has not been investigated to assess the performance of standard estimators. Andersson et al. (2013) provided the following quote: "To the best of our knowledge the operations management and revenue management literatures have not explored the use of maximum entropy methods to approximate unknown demand or willingness-to-pay distributions."

In Sheu (2010), entropy-based data fusion methodology was introduced. Sheu (2010) applied an entropy formulation to data by considering "confidence bands" that are located at one standard deviation on either side of the mean, and then between one and two standard deviations, and finally beyond three standard deviations. However, no comparative analysis was presented to demonstrate how superior or deficient the methodology may be, compared to standard estimation procedures. Grafstrom (2010) compared distributions based on entropy and suggested that the distribution with the largest entropy is more robust. In essence, if one is required to assign probabilities to events and there is no compelling justification to assign one outcome as being more likely to occur, then the events should be assigned equal probabilities which increases the entropy. Figure 1 illustrates that the entropy-based models are actually a subset of the probability based models. Decision theoretic model may or may not be constructed with a stochastic component.

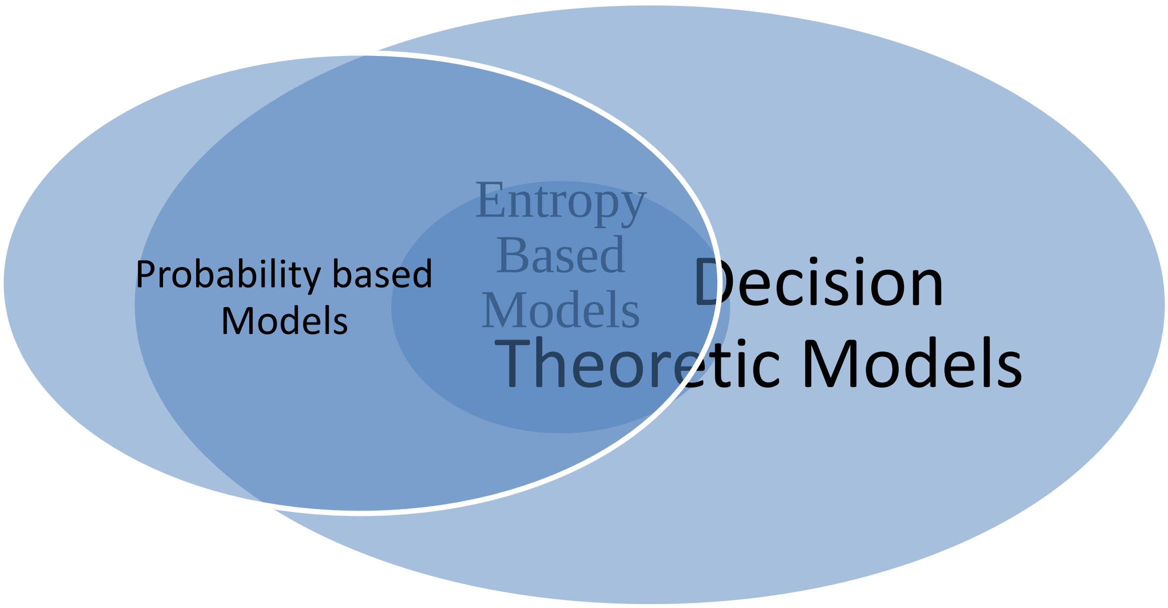


Figure 1. Entropy based models can be considered to be a subset of probability based models.

1.4. Statement of Purpose

An examination of the SCM literature reveals a significant problem: the current literature lacks an entropy-based data-fusion (EBDF) framework. This dissertation is motivated by the following problem statement:

Although standard and entropy-based methods to integrate data from multiple sources have been implemented in various applications, such as estimating the population mean or estimating distributional parameters in the newsvendor problems under limited distributional knowledge, systematic development of alternative methods and comparative studies on their performance under various conditions in SCM applications are lacking.

The purpose of this research is to establish a framework for using EBDF methods under certain conditions that may arise in SCM applications. A comparative simulation analysis employing a Monte Carlo simulation will provide a basis to investigate the robustness of the proposed EBDF methods.

## 1.5. Essays Comprising Research

This dissertation is comprised of three interrelated synergistic essays. Chapter 1 provides the introduction, motivation, research statement, practical applications of data fusion entropy methods to SCM, and research questions. The three essays are listed in Figure 2 and illustrate how the EBDF methods are introduced, assessed for estimation, and compared for performance in a newsvendor model application.

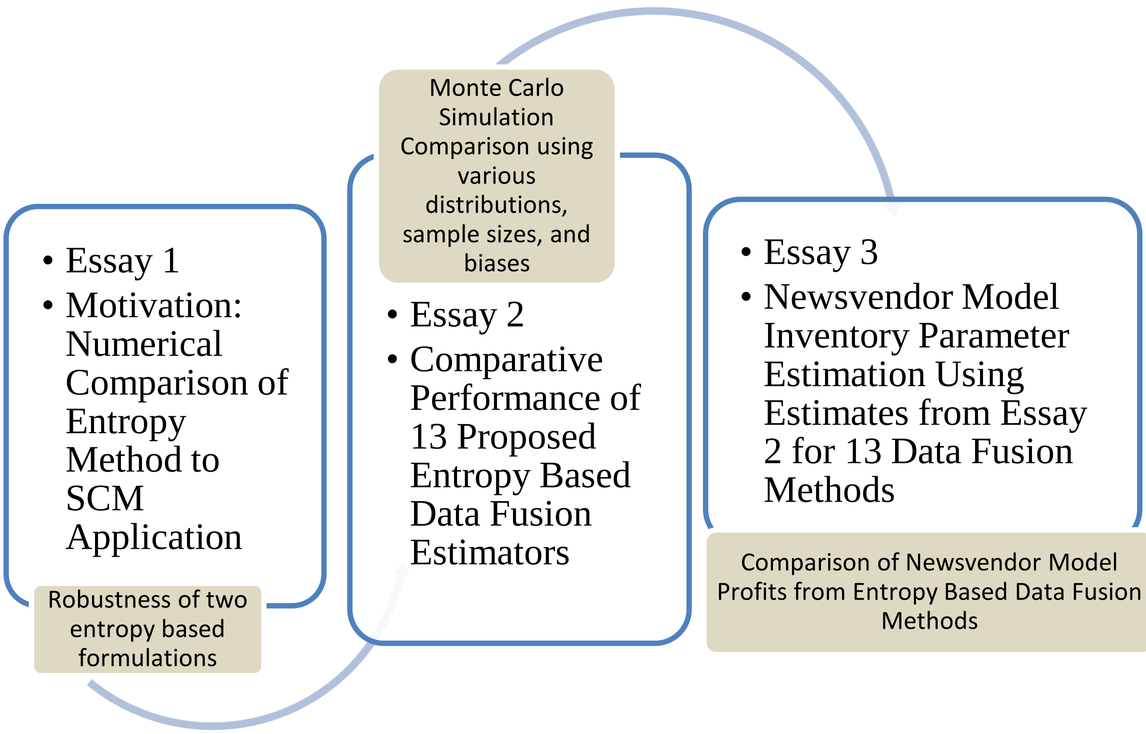


Figure 2. Layout of dissertation essays.

This section explains the research questions for each of the three separate but unified essays. A discussion is provided on how each essay addresses the research questions for this dissertation and the applicability of the results to SCM in theory and practice. Essay 1 provides an overview of supporting literature for the research question. In addition, Essay 1 demonstrates that entropy based data fusion methods may enhance decision making in an SCM application.

### 1.5.1. Essay 1 Research Question

Why should one consider alternative entropy-based weighting methods?

A numerical analysis of airline wait time data illustrates the insight that an EBDF method can provide to decision makers when using time sensitive data from several "sources." A numerical comparison of two formulations of entropy on the resulting weights from similar distributions demonstrates a compelling reason to conduct further research to assess the performance of various EBDF methods. Essay 1 provides literature support for processing data from multiple sources, which often will vary in reliability. Hence, a need exists to weigh the various sources. Obtaining reliable information is a challenge when contamination is easily acquired from biases, censorships, heterogeneity of data distributions, and inherent error from within the sources. The time sensitiveness of data in many SCM applications often does not allow for large samples of data to be collected.

How does Essay 1 research question support Problem Statement and SCM applications? By illustrating the possible robustness of two entropy based formulations, Essay 1 research question provides the motivation to examine EBDF methods and thus supports this dissertation's Problem Statement. This numerical example demonstrates how EBDF methods could assist with real-world SCM issues such as staff planning.

Essay 2 introduces thirteen data fusion methods. Ten of these methods are entropy based methods based on Sheu (2010)'s data fusion approach. An addition entropy method is included as it is based on the maximum entropy principle. A Monte Carlo simulation study examines the performance of thirteen methods in estimating the mean parameter of a population with either a normal or lognormal population distribution.

### 1.5.2. Essay 2 Research Questions

1. Can an alternative formulation to Shannon's entropy enhance the performance of Sheu's (2010)'s data fusion approach?
2. Do symmetric and skewed distributions affect the performance of these thirteen data fusion methods?
3. Do negative and positive biases affect the performance of these thirteen methods?
4. Do entropy based data fusion methods outperform non-entropy based data fusion methods?
5. Which data fusion methods are recommended for symmetric and skewed data sets when no bias is present?
6. What is the recommendation under the condition of few data sources with bias?

#### 1.5.2.1. How Essay 2 Research Question Supports Problem Statement and Applicability to SCM

By proposing and comparing the estimation performance of 11 entropy based methods and 2 traditional methods for data fusion through a Monte Carlo simulation study, this essay provides an examination of the results that leads to recommendations under various distributional and biased induced conditions that often exist in SCM applications. The research questions addressed in this essay allow for the EBDF methods and traditional methods to be contrasted for conclusions about their robustness. Thus, Essay 2 research questions provide insight into a systematic development and comparative study of the proposed data fusion methods examined in this dissertation study.

#### 1.5.2.2. How Essay 3 Research Question Supports Problem Statement and Applicability to SCM

Essay 3 explores the use of the data fusion estimates of the population mean in a newsvendor application. A Monte Carlo simulation study investigates the accuracy of the using the estimates provided in Essay 2 as the parameter estimate for the exponentially distributed demand.

### 1.5.3. Essay 3 Research Questions

1. Do data fusion methods with relatively strong performance in estimating the parameter estimate also provide relatively strong performance in estimating the optimal demand under a given ratio of overage and underage costs?
2. Do any of the data fusion methods deteriorate or improve on a relative basis with the introduction of positive and negative bias?
3. Does the alternative entropy formulations to Shannon's entropy enhance the performance of the methods on a relative basis?
4. Is the relative rank order performance of the data fusion methods different in Essay 2 and Essay 3?

#### 1.5.3.1. How Essay 3 Research Question Supports Problem Statement and Applicability to SCM

By comparing the optimal profit estimations in the newsvendor model using the inputs of a total of 13 data fusion methods in the simulation study, this essay provides an analysis of the results that lead to both theoretical and practical recommendations of appropriate methods that are robust under various distributional and biased induced conditions. Thus, Essay 3 research questions provide insight into a systematic comparative study of the proposed data fusion methods in an SCM application, namely, the newsvendor model. Essay 3 illustrates that the data fusion methods may play an important role in estimating parameters of the distribution used in the newsvendor model.

## 1.6. Methodology to Investigate Research Questions

The methodology in this research flows from the first essay to the third essay. Essay 1 uses numerical illustrations to promote the use of entropy based data fusion methods. In particular, Essay 1 provides an alternative entropy formulation and demonstrates a rationale for introducing this formulation to compute weights in data fusion methods. These methods are applied to the airport maximum wait time problem which is a proxy for SCM dynamic staff planning.

The methodology in the second essay is to provide an illustration of a framework of configurations that should be considered in assessing the robustness of data fusion methods. The normal distribution and the lognormal distribution were selected as the population distributions for the inputs of the sources and were used in the generation of the data for the Monte Carlo simulation. These distributions were selected since they are examples of symmetric and skewed distributions. The number of sources were varied to gauge the sensitivity of the methods. The number of inputs per source were varied to be small and to be moderate as might occur in a time sensitive data collection SCM application. The mean of the general population is fixed and the relative performance of the proposed data fusion methods are assessed. The reliability of the sources was controlled for by assigning various standard deviations. Larger standard deviations were assigned to less reliable sources.

The methodology in the third essay is to assess the performance of the estimation obtained by using the data fusion methods proposed in Essay 2 for the newsvendor model. The estimates of the population parameter for the newsvendor was estimated using these methods so that the optimal inventory levels could be approximated to optimize profit. The overage cost divided by the sum of the overage cost and the underage cost was set to be near one. The purpose in using this figure was to create an environment in which profit is difficult to optimize. All results from the proposed data fusion methods are compared and recommendations are provided.

The research questions in this dissertation are supported through the use of a Monte Carlo simulation analysis and a Tukey multiple comparison procedure to determine the significance of the methods. The 56 configurations in this simulation study are presented in Essay 2. Each configuration is replicated 1000 times. Simulation experiments in SCM abound in the literature since a simulation study allows multiple scenarios to be generated for analysis without a costly physical collection of data (Bachmann et al., 2013; Andersson et al., 2013; Gan & Wirth, 2005; Sheu, 2007).

The motivations for using a Monte Carlo simulation study are the following:

- True values of population parameters and optimal profit in newsvendor model are known and allow for accurate assessment of the performance of proposed methods.
- Data to assess the methods are generated from 56 configurations having various distributions with specified biases.
- Alternative methodology such as analytical analysis using mathematical derivations are not tractable in assessing the data fusion methods. That is, closed-form solutions are not available and many do not exists.
- Similar research questions have been addressed in the literature using a Monte Carlo simulation approach.

After results are available through the simulation, the proposed methods' estimates are compared using the mean absolute percent error (MAPE), the criteria explained in Essay 2. Differences in the MAPEs for estimating the mean parameter in Essay 2 and the profit for the newsvendor model in Essay 3 are analyzed for significant differences using the Tukey procedure. A distinct advantage of the Tukey multiple comparison procedure is that it controls for the family wise error rate. The family wise error is the probability of making one or more false discoveries, or type I errors, among all the hypotheses when multiple comparisons are used. An example of the use of the Tukey procedure is illustrated in Danese et al. (2013) in which pairwise comparisons of construct estimates from various countries to examine the impact of supply chain integration on responsiveness are considered.

## 1.7. Theoretical and Practical Contributions, and Summary of Conclusions from Research

### 1.7.1. Theoretical Contributions

This research proposes an extension of methodology used in Sheu (2010)'s study in modelling relief-demand management. An important contribution of the research is the introduction of Pal & Pal (1991)'s entropy formulation as an alternative approach to Shannon (1948)'s entropy definition in weighting sources of information. Another contribution of the research is the illustration on how a framework can be used to assess data fusion methodology through a Monte Carlo simulation. The obtained results provide patterns of performance by 13 proposed data fusion methods that can be partially explained using the methodological characteristics of the techniques. The overall theoretical contribution is that alternative approaches to Sheu (2010)'s data fusion approach may enhance the estimation performance of the methods under specific configurations. These alternative approaches are viable competitors with compelling results to justify their usage in data fusion applications. Further details to this research's theoretical contributions are discussed in section under each essay's summary of conclusions. The framework for data fusion methods illustrated in this dissertation may assist researchers in assessing further extensions of these methods.

### 1.7.2. Practical Contributions

This research proposed data fusion methods alternative to the one initially introduced by Sheu (2010)'s study to model relief-demand management and to facilitate emergency operations. This research highlights the importance of selecting an appropriate data fusion method in an SCM application. Knowing the number of sources, sample sizes from the sources, shape of the distribution for data inputs, and possible level of contamination of the data source would provide a practitioner with justification for selecting a data fusion methodology. The proposed data fusion methods were assessed using the newsvendor model, which has diverse applications in SCM decision making, to determine conditions in which the methods were robust. As noted in Table 1.3, numerous SCM applications may benefit from using entropy based methods: forecasting demand, personnel staffing, assessing supplier risks, simulating and modeling emergency relief efforts, and optimizing business processes and online logistical support. SCM applications requiring time-sensitive estimation under conditions in which information is limited provide attractive argument for the justification of employing an EBDF method.

### 1.7.3. Essay 1 Summary of Conclusions

Essay 1 contributes to the academic literature by presenting an alternative to Shannon's entropy formulation to compute weights for sources of information and by confirming that entropy based approaches can provide a more consistent and useful estimate even in the presence of bias. Essay 1 contributes to SCM practice by providing an SCM application in which the Transportation Security Authority at the Atlanta airport is required to track travelers' maximum wait time during each hour of the day as a quality measure and to make staffing allocation decisions at various sections of the airport. The average maximum wait time estimate must be compiled weekly for management purposes. The difficulty is its distribution is not homogeneous of the days of the week. A numerical comparison of the entropy weights for two entropy approaches, namely Shannon (1948) and Pal & Pal (1991), motivates an investigation into contrasting the performance of these approaches in data fusion applications. The results of Essay 1 support the objective of introducing the 13 data fusion methods used in Essay 2 and the establishing a framework for the use of entropy-based data fusion techniques in SCM. In addition, this essay illustrates that entropy based data fusion methods can be readily applied to decision making problems in SCM applications.

#### 1.7.4. Essay 2 Summary of Conclusions

Essay 2 contributes to the academic literature by presenting new EBDF methods and illustrating how a framework can be constructed to evaluate the robustness of the methods. In a Monte Carlo simulation study to estimate the mean of a population, the performance of the 13 data fusion methods under various distributions, biases, number of sources, and sample size (source size) is compared. For a symmetric distribution, like the normal distribution, Sheu (2012)'s method in which Shannon's entropy formulation is used is competitive in the presence of bias. However, if no bias is present, then a traditional non-entropy based approach in which the inverse variance is used is the top performing method. For a skewed distribution such as the lognormal, procedures using Pal & Pal (1991)'s entropy formulation and using intervals in the methodology consisting of a fixed size or equal to the standard deviation are among the top performing methods.

Sheu (2010)'s method with weights provided by Shannon's entropy formulation clearly performs better with inputs from a normal distribution and may perform quite poorly in the presence of many inputs from a lognormal distribution. Several procedures change rank ordering of their performance across configurations.

For many of the procedures, a positive or negative bias does not change the rank ordering performance of a data fusion method. However, a generalized maximum entropy method was shown to be easily affected by the bias and its performance depends on whether the bias is positive or negative.  The proposed entropy based data fusion methods perform better than the traditional methods when bias is present. However, in the presence of no bias, the traditional methods should be used. The traditional method in which weights are assigned by the inverse of the variance is the top performer in all scenarios without bias. In the scenario with no bias and few sources and inputs, Shannon's entropy formulation with interval either of fixed length or else equal to the overall standard deviation is recommended. Specific performance comparisons are further discussed in Essay 2.

### 1.7.5. Essay 3 Summary of Conclusions

Essay 3 contributes to the academic literature by evaluating new EBDF methods using a newsvendor model and illustrating how a framework can be constructed to evaluate the robustness of the methods in this SCM application. This essay provides an analysis of the results for the newsvendor model that leads to both theoretical and practical recommendations of appropriate methods that are robust under various distributional and biased induced conditions. The contribution of Essay 3 research to the academic literature and practitioner is the insight into the results of a systematic comparative study of the methods in the newsvendor model.  One finding is that, in general, the top performers in estimating the mean in Essay 2 repeat that performance in Essay 3 in estimating the optimal inventory level. However, there are some exceptions.

The data fusion method most affected by the type of bias is the generalized maximum entropy method similar to the case when the data fusion methods were estimating the mean of a population. Many methods employing Pal & Pal entropy perform at least as well as their Shannon entropy counterparts. As in Essay 2, a bias with large magnitude combining with a large number of sources assuming a lognormal distribution causes a strong deterioration in the method employing the inverse variance weight. More details regarding the contribution and findings of Essay 3 are provided within the essay. Essay 3 illustrates how a framework can be constructed to evaluate the robustness of the method under an important SCM application and demonstrates that the proposed methods are viable alternatives to the standard estimation procedures.

## 1.8. Overview of Dissertation Organization

This dissertation is organized into four chapters. The current chapter, Chapter 1, is an introduction and is intended to provide background and literature review regarding data fusion methodology. This chapter is an overview of the entire dissertation with research statement and questions as well as contributions and findings.

Chapter 2 presents Essay 1 in which numerical illustrations are presented to support the assertion that EBDF method are viable techniques supporting SCM decision making. This essay provides the motivation for alternative entropy based formulation in the EBDF methodology.

Chapter 3 presents Essay 2 in which 13 data fusion techniques are presented and assessed for their robustness as estimators through a Monte Carlo simulation study. This chapter provides findings and contributions to SCM academic literature by illustrating the construction of a framework that allows practitioners to select appropriate methods under various conditions for data distributions and contamination.

Chapter 4 presents Essay 3 in which the methodology assessed in Essay 2 is applied to a particular situation in the newsvendor problem. The estimates obtained from the data fusion methods in Essay 2 are used as input to the newsvendor model to optimize profit through inventory management. This essay also illustrates the construction of a framework to assess the robustness of the various methods in an SCM application. Conclusions and findings are discussed to inform practitioner how to select a robust methodology for an appropriate scenario. Chapter 4 also discusses limitations of this study. Two definitions of entropy are used in this examination of data fusion techniques. Other definitions could have been considered. The simulation study conducted in this research is not intended to be exhaustive. Patterns in the results from this simulation may not generalize to other conditions. This simulation study was conducted under a set of assumptions.

# CHAPTER 2

# ESSAY 1: NUMERICAL COMPARISON OF DATA FUSION METHODS USING TWO ENTROPY BASED METHODS

## 2.1. Abstract

This essay examines two definitions of entropy that could be used in Sheu (2010)'s data fusion weighting methodology. The literature has presented Shannon (1948)'s entropy as the standard entropy definition over many decades. Considering the long standing popularity of Shannon's formula, Pal and Pal (1991)'s entropy definition is relatively new. To integrate information from multiple sources, a data fusion technique must determine the weights of the data sources as described in Sheu (2010). This essay presents two numerical examples to contrast the computed weights based on both definitions of entropy. In the first example, this research examines real time data from the airline industry and analyzes the differences in the entropy-based data fusion estimations with respect to possible bias in the data. This numerical example demonstrates how EBDF methods could assist with real-world SCM issues such as staff planning. In the second example, probabilities representing several distributions are presented and the resulting weights provide insight into the data fusion application of Shannon's and Pal & Pal's formulations of entropy. The purpose of this essay is to motivate the use of Pal & Pal's modified definition of entropy with the ultimate objective of establishing a framework for the use of entropy-based data fusion techniques in SCM. This essay, Essay 1, provides a foundation for Essays 2 and 3 to research the robustness of proposed entropy-based data fusion techniques.

## 2.2. Introduction to Entropy-based Methodology

This essay illustrates the implementations of two entropy-based data fusion (EBDF) methods by introducing a real world application that employs these methods with entropy weights formulated by Sheu (2010): one method uses entropies formulated by Shannon (1948) and the other uses entropies formulated by Pal and Pal (1991). The objective of the application is to estimate the average of the weekly maximum wait time of travelers at a major airport; this estimate of wait time is useful in labor allocation decision making. The estimators demonstrate the sensitivity of the methods to distortion and bias in the data. The example illustrates the need to conduct a simulation study to systematically compare accuracy and variability of EBDF methods under various distributional conditions. First, the essay presents supporting literature to explain the use of EBDF in SCM applications.

The reliability of forecasts for SCM depends on accurate data and robust methodologies. Robust methodologies can overcome issues with data collected from multiple sources, data contaminated with biases, and data censored by collection constraints. Entropy-based methodology has been proposed as such a robust methodology. One of the most important concepts in information theory is Shannon’s entropy. Shannon (1948) published "A Mathematical Theory of Communication" and co-authored topics on entropy with Weaver to quantify information. Shannon defined entropy as the uncertainty in the information source which increases with the source’s randomness (Shannon, 1948; Shannon and Weaver, 1963). Shannon entropy represents the expected value of the information gained in the received messages based on the concept that the rarer the event, the more information it offers.

In various SCM applications in which data originates from multiple sources, weights may be assigned to sources in a data fusion technique to represent the corresponding reliabilities. The result is a weighted data aggregation that “fuses” or “integrates” data sources to provide an estimate that increases accuracy and lowers variability. Sheu (2010) proposed an entropy-based weighting technique in which sources with higher entropies (i.e., more uncertainties) receive smaller weights.

Sheu also examined a probability distribution based on belief levels determined through the standard deviation distance of the observation from the mean.

## 2.3. The Role of Data Fusion in SCM Applications

Since the supplier selection is a complex multi-criteria problem including both quantitative and qualitative conflicting and uncertain factors, the entropy-related research to select suppliers or logistics providers often uses multi-criteria decision analysis methods such as ELECTRE, “Elimination and Choice Expressing Reality,” (Roy, 1968), TOPSIS, “Technique for Order of Preference by Similarity to Ideal Solution,” (Yoon and Hwang, 1995), or VIKOR, “Multi-criteria Optimization and Compromise Solution,” (Opricovic, 1998) combined with a structured technique for organizing and analyzing complex decisions such as AHP, “Analytic Hierarchy Process,” (Saaty, 2000). Shannon entropy is applied to calculate objective weights for corresponding criteria. Models to select third party logistics supplier based on AHP and entropy were developed by Guoyi and Xiaohua (2011) and by Zhang et al. (2012). Lim and Shanthikumar (2007) considered a relative entropy measure for dynamic revenue management. They emphasized that the accuracy of the underlying demand rate model may not be accurately calibrated in real-world processes. Grafstrom (2010) compared distributions based on entropy and suggested that the distribution with larger entropy tend to be more robust.

Xiu and Chen (2012), and Chen and Freeman (2014) proposed using AHP, entropy weight, and TOPSIS to select third party logistics supplier and green supplier within comprehensive index systems that allow for objective and subjective weights simultaneously. Liu and Zhang (2011) constructed an indicator system with multiple criteria and combined entropy weight and ELECTRE to rank suppliers based on the net advantage value of each project. Shemshadi et al. (2011), and

Wu and Liu (2011) combined the VIKOR method and Shannon entropy concept in their models to solve multiple criteria decision problems and to deploy objective weights to indicators. Implementing Shannon entropy and linear programming, Ghorbani et al. (2012) proposed a supplier selection and order allocation model according to an analysis based on four criteria: strengths, weaknesses, opportunities and threats.

### 2.3.1. Assessing System Risk

Entropy, as a measure of uncertainty, has been used to assess the risk impact on the supply chain system. Li et al. (2010) combined fuzzy weight and entropy weight to provide an estimate of risk impact on the supply chain system and facilitate the search of system vulnerabilities. Dekkers et al. (2012) proposed using entropy coupled with simulation to assess the impact of information disruption introduced by different sources and to investigate the impact of the resulting disruption on collaborating members of the supply chain. Arkhipov and Ivanov (2011) adopted an entropy-based approach to simultaneously analyze the structural complexity and adaptation potential of the supply chain. This method can be used to select the supply chain configuration. Using fuzzy entropy, Zhang and Xu (2009) quantified the complexity of an industrial network to measure flows of goods and interaction costs between different sectors within the supply chain. Hu et al. (2008), and Isik (2010) proposed various techniques to measure and mitigate complexity to assist in designing systems with robust performances. Wang et al. (2005) applied Shannon's entropy and entropic indices to measure the complexity of dynamic decision processes and of the Markov decision processes under random and deterministic policies. Gan and Wirth (2005) combined an empirical approach and an entropy measure to determine the validity of the transition between deterministic scheduling and online scheduling. Jiang et al. (2012) found that the entropy analysis of technology standardization clarified the interrelationship between technology standards and industrial innovations.

Since entropy is a measurement for disorder, Scholz-Reiter et al. (2007) proposed using entropy to measure the quality of demand forecasting. Isik (2011) proposed a complexity management model which covers identifying, measuring, analyzing and controlling complexity since reducing complexity is reducing high costs to the company. To measure the firm's capability to quickly respond to customer demands, Shuiabiet al. (2005) proposed using entropy to measure and monitor the flexibility of manufacturing operations. Huatuco and Shaw (2010) demonstrated that the robustness of supply chains can be assessed using entropic-related complexity retrospectively and prospectively. Sivadasan et al. (2002) described a technique to measure the operational complexity of the supply chain based on entropy to quantitatively detect and prioritize operational complexity hotspots.

Using simulation and entropy, Martínez-Olvera (2008) created an entropy-based formulation to compare different information sharing approaches in a supply chain. Allesina et al. (2010) developed indexes to measure a supply network complexity and solve the problem of supply network optimization based on entropy of information.

### 2.3.2. Forecasting Measures of Logistics

Goh and Law (2003), and Goh et al. (2008) analyzed tourism demand based on rough sets theory and minimal entropy partitioning algorithm. In a large-scale disaster, given various estimations of fatalities produced by multi-sensors (different information sources) with multiple estimations per sensor, Sheu (2010) used a team consensus approach to estimate the dynamic fatalities. In this method, the entropies associated with the information sources are used to calculate weights to linearly combine the means of the sensors' estimates. The weights reflect the relative reliability associated with these sources of information: sources with higher entropies (i.e., more uncertainties) receive smaller weights. Cheng et al. (2006) proposed an approach to forecast IT project cost through combining minimize entropy principle approach (MEPA) and fuzzy time series forecasting methods.

### 2.3.3. Enhancing Systems

To reduce cost and to gain improvements to production systems, Jaber and Rosen (2008) suggested various approaches to reduce system entropy in production systems. Jaber et al. (2014) showed that the performance of the consignment cost, a system in which vendors stock inventory at the buyer's location, improve when entropy is reduced. Jaber et al. (2006) and Jaber (2007) enhanced the lot size model, which is the economic order quantity model, by reducing the hidden cost through reducing system cost entropy.

In the airline industry, researchers proposed using the maximum entropy approach to update the booking limits under curbed demand information. This approach utilizes past observations to obtain a discrete probability distribution which can be used to determine the booking limits (Eren and Maglaras, 2006; Lan et al., 2008).

### 2.3.4. Improving the Newsvendor Model's Inventory Estimation

Andersson (2013) considered the newsvendor model under partial information and used a maximum entropy approach to assess the performance of standard estimators and to generate the most likely distributions. Perakis and Roels (2008) generated criteria to select the demand distribution as an input to the newsvendor model using the maximum entropy approach. Saghafian and Tomlin (2014) proposed a non-parametric, maximum-entropy based technique to combine demand observations with tail-behavior of distribution functions which can dynamically respond to hidden changes in the unknown true distribution.

### 2.3.5. Summary of Entropy Application for SCM Models

Table 2.1 depicts important SCM applications that may benefit from entropy methodology. These contributions from the literature are grouped in eleven categories that appear frequently. Several researchers contribute to several of these areas. These papers help to support the investigation of entropy concepts in SCM applications.

Table 2.1 Summary contributions of entropy based methods in SCM

| | Assessment | Improvement of Business | Forecasting | Shannon | Minimum Entropy | Maximum Entropy | Generating Distribution | Newsvendor | Multi-Criteria Ranking | Multi-Criteria Grouping | Simulation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Allesina et al. (2010) | X | X | | X | | | | | | | |
| Andersson et al. (2013) | | | | | | X | X | X | | | |
| Arkhipov and Ivanov (2011) | X | X | | X | | | | | | | |
| Chen and Freeman (2014) | X | | | X | | | | | X | | |
| Cheng et al. (2006) | | | X | | X | | | | | X | |
| Dekkers et al. (2012) | X | | | X | | | | | | | |
| Eren and Maglaras (2006) | | | X | | | X | X | | | | |
| Gan and Wirth (2005) | X | | | X | | | | | | | |
| Ghorbani et al. (2012) | X | | | X | | | | | X | | |
| Goh et al. (2008) | | | X | | X | | | | | X | |
| Goh and Law (2003) | | | X | | X | | | | | X | |
| Guoyi and Xiaohua (2011) | X | | | X | | | | | X | | |
| Hu et al. (2008) | X | X | | X | | | | | | | |
| Huatuco and Shaw (2010) | X | X | | X | | | | | | | |
| Isik (2010) | X | X | | X | | | | | | | |
| Isik (2011) | | X | | X | | | | | | | |
| Jaber et al. (2006) | | X | | X | | | | | | | |
| Jaber et al. (2007) | | X | | X | | | | | | | |
| Jaber et al. (2014) | | X | | X | | | | | | | |

| | Assessment | Improvement of Business | Forecasting | Shannon | Minimum Entropy | Maximum Entropy | Generating Distribution | Newsvendor | Multi-Criteria Ranking | Multi-Criteria Grouping | Simulation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jaber and Rosen (2008) | | X | | X | | | | | | | |
| Jiang et al. (2012) | | X | | X | | | | | | | |
| Lan et al. (2008) | | | X | | | X | X | | | | |
| Li et al. (2010) | X | | | X | | | | | | X | |
| Liu and Zhang (2011) | X | | | X | | | | | X | | |
| Martínez-Olvera (2008) | X | X | | X | | | | | | | X |
| Perakis and Roels (2008) | | | | | | X | X | X | | | |
| Saghafian and Tomlin (2014) | | | | | | X | X | X | | | |
| Shemshadi et al. (2011) | X | | | X | | | | | X | | |
| Sheu (2010) | | | X | X | | | | | | X | |
| Scholz-Reiter et al. (2007) | X | | X | X | | | | | | | |
| Shuiabiet al. (2005) | X | X | | X | | | | | | | |
| Sivadasan et al. (2002) | X | X | | X | | | | | | | |
| Wang et al. (2005) | X | | | X | | | | | | | |
| Wu and Liu (2011) | X | | | X | | | | | X | | |
| Xiu and Chen (2012) | X | | | X | | | | | X | | |
| Zhang et al. (2012) | X | | | X | | | | | X | | |
| Zhang and Xu (2009) | X | | | X | | | | | | X | |

## 2.4. Importance of Entropy-Based Data-Fusion (EBDF) Framework in SCM

Effective forecasting impacts the efficiency of supply chains and affects the management of operations, logistics, and merchandising. Effective forecasting relies on accurate data and robust methodologies. However, multiple issues causing data inaccuracy or conflict affecting many SCM applications exist, including but not limited to data coming from multiple sources, data being biased, or data being censored. Entropy-based methodology has been introduced to overcome these data issues. However, no comparative analysis was presented to demonstrate how superior or deficient these methods may be compared to standard estimation procedures.

### 2.4.1. Data Issues in SCM

Poor quality data affects the management of operations, logistics, and merchandising since it distorts demand forecast, and causes the "bullwhip" effect that leads to unreliable deliveries, high safety stock and subsequent stock shortages (Lee et al., 1995). Poor data also shrinks operating revenue and share-price (Hendricks and Singhal, 2005), and obstructs risk management (Hendricks and Singhal, 2005; Curkovic, 2015). Poor data may also have deadly consequences. The Institute of Medicine estimated that medical data errors kill between 44,000 and 98,000 people a year in US hospitals (English, 2009). The most recent Ebola outbreak was one of the worst in history because of data problems. According to the Center of Disease Control , the underreporting of Ebola cases resulted in the rescue efforts, in terms of supplies and manpower, being delayed three or four months (Voelker, 2014). Data problems reduce profit, alienate customers, and hinder new strategies (Redman, 1995). Even so, problems with data exists for many reasons: data gathered from multiple sources, bias introduced intentionally or unintentionally into the data, and data recorded incompletely or censored.

### 2.4.2. Multiple Sources in SCM Operations

The exponential growth of data availability through digital technology development has increased SCM's reliance on information as a strategic resource (Ballou et al., 2003). However, as the variety and volume of the collected data increases, data quality suffers (Coyle et al., 2012). Since supply chains span many organizations, the data quality problem is compounded by the reliability of collection methods, collection frequency, and multiple collection sources (Harrison et al., 2005; Lee and Strong, 2003). For example, emergency management programs designed to respond to crises such as earthquakes and hurricanes, often relies on data from armed forces, governments, private organizations, or humanitarian agencies, each of which employs its own method of operations (Schulz & Blecken, 2010; Davis, 2010). In emergency management situations, data must be "fused" from several sources in a time-sensitive manner. Consequently, decision-making questions are addressed by integrating multiple heterogeneous data information sources (Madnick et al., 2003).

### 2.4.3. Data Bias and Distortion

SCM data bias can be categorized as either intentional or unintentional and is a major threat to effective decision making. Intentional bias may occur in demand forecasting through personal motivations or misalignment of incentives. For example, a sales department manager may inflate the forecasted data to guarantee product availability. This availability may allow for an increase in sales commissions. On the other hand, an operations manager may smooth the forecasts to avoid costly production swings (Shapiro, 1977; Oliva and Watson, 2009). Unintentional bias is an unforeseen and uncontrollable error, often caused by process deficiencies, lack of information, or lack of management experience in forecasting methodology (Makridakis et al., 1998). An example of unintentional bias is the over-estimation of sales due to the lack of experience of a new manager. Intentional and unintentional biases stymie forecast accuracy, distort demand information, and cause negative supply chain phenomenon such as the "bull-whip" effect (Lee et al., 1992).

### 2.4.4. Censored Data

Data censoring is a condition in which observations are censored, meaning that their values are recorded correctly only if they are below or above a specified value. This condition may happen when demand is not recorded after demand is satisfied. Data could be censored from above (right-censored) or below (left-censored) and may bias final estimates of interest. As an example of censored from above data, Greene (2003) suggests the example of ticket sales to sporting events, in which the actual latent demand is recorded correctly only if the ticket sales are below the facility capacity. After the events are sold out or reach capacity, the demand is not recorded. The recorded data, incomplete and censored at the ticket limit, reflects the ticket sales, not the true demand. Data censorship is particularly complex among businesses practicing revenue management such as airlines since there are multiple fare classes subject to a common capacity constraint. Demand forecasting is difficult with censored data since the information is incomplete and often yields forecasts carrying a negative bias that underestimates the true demand (McGill, 1995).

## 2.5. Probabilistic Formulation of Entropy Based Data Fusion

Shannon's definition of entropy, the most widely used formulation for entropy, is defined as the expectation of the information gain $I(x)$. For each event $x_i$, $I(x_i)$ is described as the negative of the logarithm of its probability. Since $I(x_i)$ can be unbounded, the information gain has no upper limit. In practice, the computation of the logarithm of a very small probability may cause computational instability.

Shannon's definition of entropy has a long history, especially in information theory. Shannon defined entropy as the expectation of the information gain $I(x)$, measured by the negative of the logarithm of the probability of an event. Notationally, if $H(x)$ represents the entropy then its computational formula is expressed as follows:

$$H(x) = E(I(x)) = \sum_i P(x_i)\, I(x_i) = -\sum_i P(x_i)\, log_b\big(P(x_i)\big) \qquad (1)$$

where

$$I(x_i) = -\, log_b\big(P(x_i)\big) = log_b\Big(\frac{1}{P(x_i)}\Big), \qquad (2)$$

$\sum_i P(x_i) = 1,$

$$\lim_{P \to 1^+} Plog_b(P) = 0, \tag{3}$$

and

$$\lim_{P \to 1^-} Plog_b(P) = 0 \tag{4}$$

Defining $I(x_i) = -\,log_b\big(P(x_i)\big) = log_b(\frac{1}{P(x_i)})$ could cause computational problems in practice since if $p(x_i) \to 0$, then $I(x_i) \to \infty$. Since $I(x_i)$ can be unbounded, Shannon entropy can also become unbounded and can be problematic in certain applications. Researchers have modified the Shannon's entropy, by redefining the information gain function $I(x)$.

Pal and Pal (1991) modified Shannon's entropy by redefining the information gain function $\mathrm{I(x)}$. They proposed a new definition of classical entropy based on the exponential behavior of information gain and suggested that the measure of "uncertainty" in information gain is better represented by $1 - \mathrm{p(x_i)}$ than by $(1/\mathrm{p(x_i)})$. Instead of $-\log(1 - \mathrm{p_i})$, the information gain is redefined as $\mathrm{I(p_i) = e^{1-p_i}}$. The entropy function becomes

$$H(x) = E(I(x)) = \sum_i P(x_i)\, I(x_i) = \sum_i P(x_i)\, e^{1-P(x_i)}$$

The new information gain function is monotonically decreasing and bounded between 1 and e as p decreases from 1 to 0. This function avoids the awkward situation of approaching infinity as $\mathrm{p(x_i)} \to 0$ .

In general, Pal and Pal's entropy is bounded between 1 and e since the information gain is bounded between 1 and e, while Shannon's entropy could become unnecessarily large. Because Pal and Pal's entropy is bounded, Pal and Pal's entropy may have less variability than Shannon's entropy. Table 2.4 illustrates differences in the computed Pal and Pal's entropy and Shannon's entropy for discrete uniform distributions.

Sheu (2010) proposed an entropy-based weighting technique based on the general principles of multi-sensor fusion and the team consensus approach (Chung and Shen, 2000) which combines data form all sources (teams) into a final approximation equaling the weighted average of all source means. The weights reflect the relative reliabilities associated with the sources (teams), meaning that sources with higher entropies (i.e., more uncertainties) receive smaller weights (Chung et al., 1997; Sheu, 2010).

Table 2.4 Comparison of Shannon's and Pal & Pal's entropies for discrete uniform

| N = Number of Data Points | $P(x_i) = 1/N$ | Shannon Entropy | Pal and Pal Entropy |
|---|---|---|---|
| 2 | 0.5 | 0.69 | 1.65 |
| 12 | 0.083333333 | 2.48 | 2.50 |
| 25 | 0.04 | 3.22 | 2.61 |
| 100 | 0.01 | 4.61 | 2.69 |
| 1000 | 0.001 | 6.91 | 2.72 |
| 10000 | 0.0001 | 9.21 | 2.72 |
| 100000 | 0.00001 | 11.51 | 2.72 |
| 1000000 | 0.000001 | 13.82 | 2.72 |

In this entropy-based weighting scheme, if K data sources and K estimated entropies $H_i$ are known, the K appropriate weights $w_i$ can be determined by solving an optimization problem to minimize the overall entropy. This is an extended classic Lagrangian problem whose objective is to minimize the sum of the squares of the weighted entropies where boundary conditions include inequalities. The problem is formulated as follow.

Minimize $\Omega = \sum_{i=1}^{K} (w_i \, H_i)^2$ , $i = 1, 2, \ldots K$

subject to the conditions:

$\sum_{i=1}^{K} w_i = 1$ and $w_i > 0$ , $i = 1, 2, \ldots K$.

With the application of the Kuhn-Tucker method, Ω is minimized if

$$w_i = \frac{1}{{H_i}^2 \sum_{i=1}^{K} {H_i}^{-2}}, \qquad i = 1, 2, \dots K$$

After $K$ weights $w_i$ are defined together with $K$ source means $u_i$, $i = 1, 2, \dots K$, the integrated forecast fusing data from $K$ sources is a weighted average of $K$ source means $u_i$, $i = 1, 2, \dots K$. The integrated forecast is $X = \sum_{i=1}^{K} w_i \ u_i$, when information from K data sources are fused.

## 2.6. Intervals for Belief-Strength Bands Used to Provide Posterior Probabilities

Sheu (2010) provided one formulation of an entropy-based data fusion method by using posterior probabilities of data belief-strength bands to account for the reliability of the collected data. Sheu (2010) assumes that a random sample $\mathrm{X}_1, \dots, \mathrm{X}_\mathrm{n}$ follows a Gaussian distribution with mean μ and standard deviation s, and lets M be the number of levels of belief strengths and $\mathrm{B}_\mathrm{m}$ be a band of data around the mean in the following manner:

$$B_m = \{ x_j \mid (\mathrm{m}-1)\mathrm{s} < |x_j - \mu| \le \mathrm{ms}, 1 \le j \le n\}, \quad \mathrm{m} = 1, 2, \dots \mathrm{M}.$$

Let $\mathrm{N_i}$ be the number of data contained in $\mathrm{B_m}$. Sheu defined the posterior probability as

$$p(m|\{x_1, x_2, \dots, x_n\}) = \frac{N_m}{n}, \quad \mathrm{m} = 1, 2, \dots \mathrm{M}$$

Hence, the entropy of the information based on Shannon's formulation is as follow:

$$H(\{x_1, x_2, \dots, x_n\}) = -\sum_{m=1}^{M} p(m|\{x_1, x_2, \dots, x_n\}) \log(p(m|\{x_1, x_2, \dots, x_n\}))$$

In other words, given n values from a data source, a distribution for M belief levels is constructed before the calculation of Shannon's entropy. First, different data belief-strength bands around the mean are specified; then the percentage of data lying within each belief-strength band is used as the posterior probability for a particular belief level. Sheu (2010) used the percentage of data lying within one standard deviation of the mean as the posterior probability for belief level 1, the percentage of data lying between one and two standard deviations from the mean as the posterior probability for belief level 2, and the percentage of data lying within at least 3 standard deviations from the mean as the posterior probability for belief level 3. The following figure illustrates the calculations of the posterior probabilities for a set of data following a normal distribution.

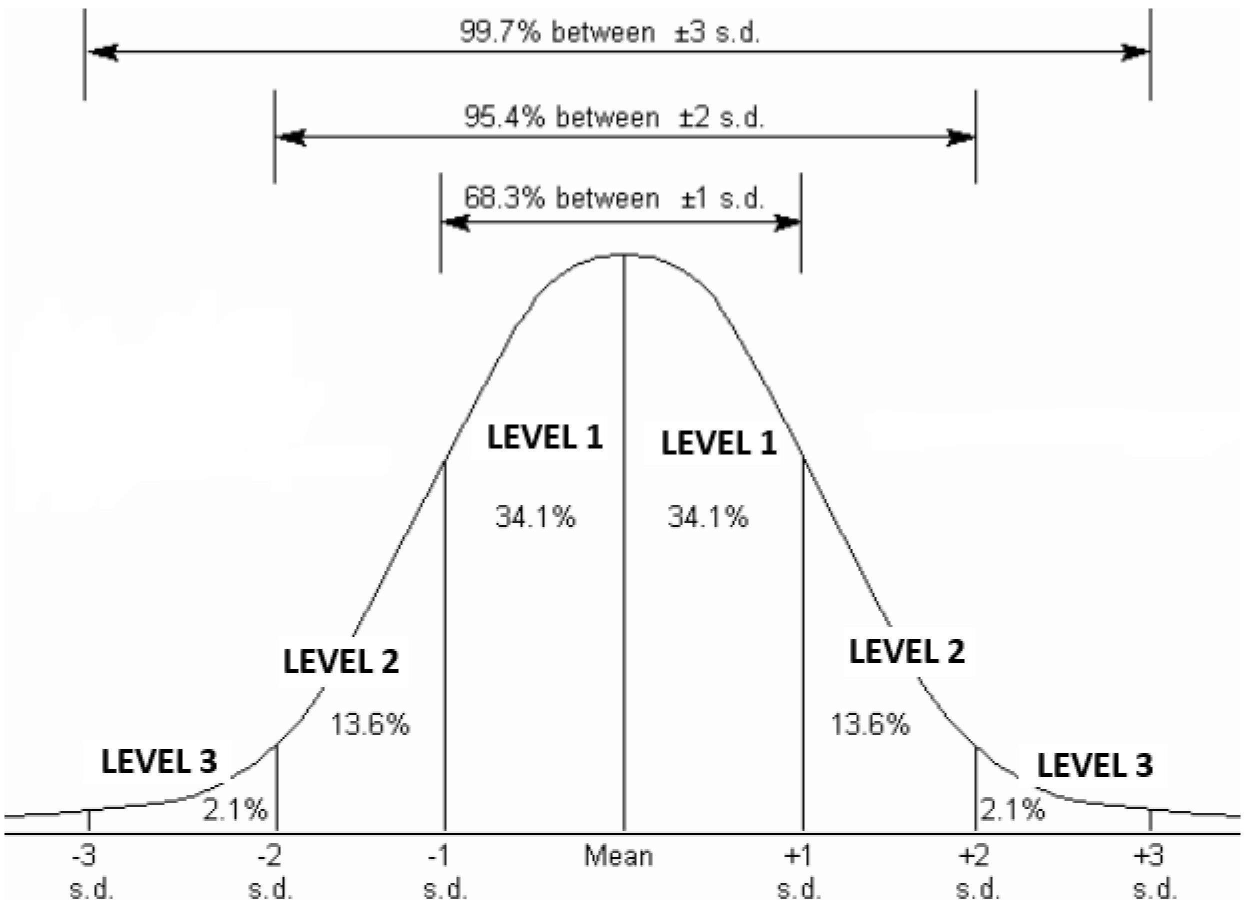


Figure 3. Distribution of levels of belief strength for a Gaussian population according to the empirical rule.

According to the empirical rule, if the data from a source follows a normal distribution, then the posterior probability for belief level 1 is 68.3%, for belief level 2 is 27.1%, and for belief level 3 is 4.6%. The empirical rule only holds for the normal distribution, but is a useful approximation for many unknown distributions in practice. Sheu's method will yield different posterior probabilities for non-Gaussian distributions. With small sample sizes, the estimates of these posterior probabilities may be easily influenced by variation in the data.

Many SCM applications require integrating data from multiple sources. For example, emergency logistic managers during a major disaster such as an earthquake or an epidemic often must base their urgent decisions on conflicting information originating from multiple information sources. The information sources in the emergency logistics context could be on-the-spot groups of reporters, rescuers, and charities, each of which contains diverse members providing diverse estimates of fatalities which must be integrated to coordinate relief efforts. The assumption used here is that data from a data source (a data source could be a group of reporters, a group of rescuers, or a group of charities) follows a certain distribution and that the sample means of these data sources are considered approximations of the population parameters. Weights may be assigned to these sample means to represent the reliability of the sources. A weighting scheme often needs to be implemented quickly as real-time data arrives so that the time-sensitive population parameters can be estimated. The result is a weighted data aggregation that "fuses" or "integrates" data sources to provide an estimate that increases accuracy while lowering variability.

## 2.7. Numerical Illustration of Decision Making with Entropy Based Data Fusion

To illustrate the importance of weighting data sources in a time-sensitive application, the following example examines a sample of wait times from Atlanta's international airport. This numerical example demonstrates how EBDF methods could assist with real-world SCM issues such as staff planning. One performance measure of manpower allocation is the average maximum wait time which could be used to allocate airport staffing to relieve the demand on airport screening, equipment, and ticketing services. A very high average maximum wait is an indication of poor manpower planning. Managerial time-sensitive decisions on staffing need to address the ever-changing challenges of flights and events. Decisions on manpower could be projected dynamically using weekly data.

The Wall Street Journal reported on July 20, 2016 that TSA Security lines were not as bad as most individual feared (http://www.wsj.com/articles/why-tsa-security-lines-arent-as-bad-as-youd-feared-1469032116). The measure that was used for performance was the maximum average wait time. The article reported the following: “Denver’s airport says it has had the 10 busiest days in its history this summer, but the average of maximum wait times for each day in June was 18 minutes, down from an average of about 25 minutes in each month between February and May. At Chicago O’Hare, passengers were stuck overnight in terminals after missing flights in May.” Obviously, when this performance measure, namely, the average maximum wait time, decreases, the TSA is praised for improving transportation security processes.

Table 2.6.1 presents the maximum wait times for various hourly time blocks over a one-week period. The data is collected from the time period May 31 to June 3, 2016. In this example, the days of the week will represent “sources” and the hourly maximum wait times represent the “inputs” within the source. Wait times tend to be heterogeneous across days of the week. Two proposed methods of approximating the weekly maximum wait time are used in this illustration along with the traditional average. This application underscores the robustness of these estimates and their sensitivity to bias. Sheu (2010)’s EBDF technique using Shannon’s entropy definition and a modification of his method using Pal and Pal (1991)’s entropy definition are compared.

The true average of the wait times is 43.24. The estimates of the average max wait time using Sheu’s weighting technique and Shannon entropy, and Sheu’s weighting technique with Pal

& Pal Entropy are 44.02 and 43.62 respectively. This agreement occurs because the distribution does not vary dramatically among the days. To illustrate the advantage of EBDF technique, biases are added to data. In particular, let's suppose that the wait times on Thursday was distorted and decreased by 20 minutes because of a systematic error, and that the wait times on Friday were increased by 50 minutes due to a lack of personnel. The new estimates of the average max wait time using Sheu's weighting technique and Shannon entropy, and Sheu's weighting technique with Pal & Pal Entropy would be 42, 48, and 50 minutes respectively.

Table 2.6.1 Max wait time in minutes at Atlanta Airport - May 31, 2016 to June 3, 2016

| Max Wait Time at Atlanta Airport | | | | | |
|---|---|---|---|---|---|
| Entering Security | Monday | Tuesday | Wednesday | Thursday | Friday |
| Day of the Week | 2 | 3 | 4 | 5 | 6 |
| 06:00 – 07:00 | 3 | 32 | 38 | 11 | 51 |
| 07:00 – 08:00 | 25 | 30 | 16 | 11 | 36 |
| 08:00 – 09:00 | 0 | 73 | 17 | 30 | 8 |
| 09:00 – 10:00 | 3 | 63 | 0 | 73 | 24 |
| 11:00 – 12:00 | 37 | 55 | 68 | 59 | 25 |
| 13:00 – 14:00 | 34 | 58 | 66 | 58 | 76 |
| 14:00 – 15:00 | 42 | 26 | 35 | 40 | 106 |
| 15:00 – 16:00 | 41 | 39 | 34 | 31 | 103 |
| 16:00 – 17:00 | 28 | 50 | 63 | 65 | 123 |
| 17:00 – 18:00 | 24 | 49 | 49 | 115 | 66 |
| 18:00 – 19:00 | 32 | 39 | 70 | 38 | 38 |
| 19:00 – 20:00 | 26 | | 44 | 11 | 40 |
| 20:00 – 21:00 | 45 | | | | 77 |
| 21:00 – 22:00 | | | | | 45 |
| 22:00 – 23:00 | | | | | 30 |

Table 2.6.2 Estimates of the average maximum wait time

| Estimate of Average Max Wait Time at Atlanta Airport | | | |
|---|---|---|---|
| Method | Entropy Weight<br><br>Sheu's with Shannon Entropy | Entropy Weight<br><br>Sheu's with Pal & Pal Entropy | Equal Weight |
| | 44.02 | 43.62 | 43.24 |

Table 2.6.3 Maximum wait time in minutes at Atlanta Airport from May 31, 2016 to June 3, 2016.

| Max Wait Time at Atlanta Airport | | | | | |
|---|---|---|---|---|---|
| Entering Security | Monday | Tuesday | Wednesday | Thursday Bias = -20 | Friday Bias = +50 |
| Day of the Week | 2 | 3 | 4 | 5 | 6 |
| 06:00 – 07:00 | 3 | 32 | 38 | 0 | 101 |
| 07:00 – 08:00 | 25 | 30 | 16 | 0 | 86 |
| 08:00 – 09:00 | 0 | 73 | 17 | 10 | 58 |
| 09:00 – 10:00 | 3 | 63 | 0 | 53 | 74 |
| 11:00 – 12:00 | 37 | 55 | 68 | 39 | 75 |
| 13:00 – 14:00 | 34 | 58 | 66 | 38 | 126 |
| 14:00 – 15:00 | 42 | 26 | 35 | 20 | 156 |
| 15:00 – 16:00 | 41 | 39 | 34 | 11 | 153 |
| 16:00 – 17:00 | 28 | 50 | 63 | 45 | 173 |
| 17:00 – 18:00 | 24 | 49 | 49 | 95 | 116 |
| 18:00 – 19:00 | 32 | 39 | 70 | 18 | 88 |
| 19:00 – 20:00 | 26 | | 44 | 0 | 90 |
| 20:00 – 21:00 | 45 | | | | 127 |
| 21:00 – 22:00 | | | | | 95 |
| 22:00 – 23:00 | | | | | 80 |

Table 2.6.4 Estimates of the average maximum wait time under data distortion.

| Estimate of Average Max Wait Time at Atlanta Airport | | | |
|---|---|---|---|
| Method | EBDF estimate Sheu's with Shannon Entropy | EBDF estimate Sheu's with Pal & Pal Entropy | Equal Weight |
| Estimate Max Wait | 42 | 48 | 50 |

These estimates illustrate how the data fusion methods are affected by the distortion. Data fusion method selection is an important part of the decision making process and decision makers need to be aware of the potential sensitivity of the method to distortion or bias in the data. In this case, Sheu's Shannon Entropy method is still somewhat close to the unbiased mean value of 43. The equal weight and Sheu with Pal and Pal have a larger bias for this example.

## 2.8. Contrasting Shannon's and Pal & Pal's Entropy Formulations

To better contrast the use of Shannon's Entropy and Pal & Pal's entropy formulations, the following probability distributions are presented for 5 groups in Table 2.7.1. Viewing these distributions across Groups 1 through 5, one could readily conclude that the distributions are similar. Even though Group 5 is somewhat different, its pattern of two very small probabilities and one large probability is consistent with the previous groups. This suggests that weights for the groups should be similar.

The next illustration presents the entropy computations using Shannon's and Pal & Pal's formulators and the corresponding weights using Sheu's method using the distributions in Table 2.7.1. The entropies using Shannon and Pal & Pal's computations are equal across the first four groups. However, the entropy for Group 5 is relatively much less using Shannon's computation compared to Pal & Pal's. This relative difference results in weights that are considerably different for the five groups. Sheu's weight for Group 5 is close to one. This weight occurs because Group 5 can be considered to have fewer "outliers" and to have relatively higher reliability than the other four groups. However, Sheu's weights using Pal & Pal's entropy results in practically equal weights as shown in Table 2.7.2. These weights can allow for the information from all groups to be considered in the calculation of unknown distribution parameters, such as the mean.

Table 2.7.1 Probability Distributions for 5 Groups

| | Probability Distributions for 5 Groups | | | |
|---|---|---|---|---|
| | Prob1 | Prob2 | Prob3 | Sum |
| Group1 | 0.003 | 0.003 | 0.994 | 1 |
| Group2 | 0.003 | 0.003 | 0.994 | 1 |
| Group3 | 0.003 | 0.003 | 0.994 | 1 |
| Group4 | 0.003 | 0.003 | 0.994 | 1 |
| Group5 | 0.0001 | 0.0001 | 0.9998 | 1 |

One could argue that Sheu's method with Shannon's entropy puts too much weight on one group, in this example, and this comes at the price of ignoring the information from the other groups.

Table 2.7.2 Contrasting Sheu's Weights with Shannon's and Pal & Pal's Entropies

| | Contrasting Sheu's Weights with Shannon and Pal & Pal Entropies | | | | |
|---|---|---|---|---|---|
| | Shannon's Entropy | Pal & Pal's Entropy | | Sheu's Weight based on Shannon's Entropy | Sheu's Weight based on Pal & Pal's Entropy |
| Group1 | 0.040837 | 1.016243 | | 0.0025 | 0.199 |
| Group2 | 0.040837 | 1.016243 | | 0.0025 | 0.199 |
| Group3 | 0.040837 | 1.016243 | | 0.0025 | 0.199 |
| Group4 | 0.040837 | 1.016243 | | 0.0025 | 0.199 |
| Group5 | 0.002042 | 1.000544 | | 0.99 | 0.204 |

This illustration is useful to understand that these estimators react to assumptions of the data and to possible bias. To recommend one of these estimators under various conditions, this research employs a simulation study to compare accuracy and variability as well as underlying bias from the forecast. The mean absolute percent error (MAPE) is used to compare the accuracy and variability of forecasting estimates since this measure is often used in setting safety stock for inventory purposes.

## 2.9. Conclusion

Essay 1 contributes to the theoretical (academic) literature by presenting numerical illustrations that reveal a need for alternative entropy weighted methods in data fusion estimations. Essay 1 contributes to SCM practice by demonstrating the effect of selecting different entropy formulations. This essay primarily explains the use of Shannon's entropy and Pal & Pal's entropy, and contrasts resulting weights in an EBDF application. The examples in the essay justify a systematic study of weighting approaches in data fusion applications. In the literature, Shannon's entropy is considered to be the standard measure of "transmitted information" or "signal meaning" in information theory. This entropy has stood the test of time as the primary measure of "uncertainty" and "diversity" since Claude Shannon introduced it in 1948. In the next couple of essays, the benefit of using this entropy measure in Sheu's weighting scheme will be systematically examined with other formulations involving distributional assumptions. The illustration contrasting Shannon's and Pal & Pal's entropy weighting techniques reveals that each procedure may have merits/shortcoming worth exploring in a systematic comparative study.

# CHAPTER 3

# ESSAY 2: PARAMETER ESTIMATION: COMPARATIVE PERFORMANCE OF PROPOSED ENTROPY AND NON-ENTROPY BASED DATA FUSION ESTIMATORS

## 3.1. Abstract

This essay presents a Monte Carlo simulation study regarding the estimation of the mean parameter of a population based on data inputs from multiple sources. Several EBDF methods are proposed. The bench mark EBDF method is Sheu (2010)'s, which was defined in Essay 1 as an application of Shannon's entropy using belief bands of data centered around the mean. Essay 1 also discussed the use of Pal & Pal (1991)'s definition of entropy, which can be used as an alternative option to Shannon's entropy in EBDF. Essay 2 presents new formulations of EBDF methods by considered different interval belief bands as alternative options to Sheu (2010)'s belief bands. This essay also examines a generalized maximum entropy method developed by SAS and two traditional estimators of the population mean: the inverse variance weighted and the ordinary average techniques. The performance of these methods will be assessed using normally distributed and lognormally distributed data generated through the simulation study. The sensitivity of these methods to bias in the data, the number of sources, and the distributions of the underlying data are examined.

## 3.2. Introduction

This section introduces the types of EBDF methods that are examined in a simulation study. The contribution of this analysis to SCM applications is to provide new EBDF methods to the academic literature and provide practical guidelines to the use of SCM data fusion applications. In today's rapid technology environment with multiple available media outlets, strategies need to be developed to integrate data in a timely fashion to enhance SCM agility. To remain competitive in a dynamic business environment, SCM needs to be responsive to changing intelligence from a variety of sources including customer demand, supplier trends, diffusion of technology, and evolving roles of social and news media. A gap exists in the literature in developing techniques for integrating diverse data for dynamic decision making. Sheu (2010) provided an innovative procedure to estimate mean survival population figures during an aftermath of a major disaster. His method incorporated an entropy based weighting of data sources to provide more robust estimation for the mean parameter of the number of fatalities using standard assumptions. However, published research assessing the performance of entropy based methods in statistical estimation are not readily available.

Essay 2 contributes to the literature by addressing the robustness of Sheu (2010)'s method as well as other proposed methods using both normally and lognormally distributed data. Sensitivity analysis of these methods with regard to positive and negative biases, the number of sources, and the distributions of the underlying data is conducted.

## 3.3. Proposed Entropy Based Data Fusion Methods

As provided in Essay 1, Sheu (2010) proposed a formulation to weight sources of information using entropy methodology. The general idea is that confidence bands or intervals of the data provide the foundation to construct a probability distribution of the belief levels to determine the reliability of the sources. This essay introduces modifications to Sheu (2010)'s computation of a posterior distribution of belief levels by allowing the confidence bands or intervals to be determined in various fashions. In addition, the use of an alternative definition of entropy is incorporated. Sheu (2010)'s method will be a bench mark to which the alternative methods are assessed.

Table 3.2.1 Proposed EBDF Methods Using Sheu (2010)’s Methodology

| Meth. | Defining Intervals for Meth. | Entropy | Rationale for Meth. |
|---|---|---|---|
| SHN_ Sheu | Three belief bands: First consists all values within one standard deviation of the mean; Second consists of values between one and two standard deviations; third consists of all values exceeding 2 standard deviations. | Shannon's Entropy | Data within each belief band has the same reliability. Proposed by Sheu (2010). Pal & Pal (1991) suggest advantages to alternative entropy formulations. |
| PAL_ Sheu | | Pal & Pal's Entropy | |
| SHN_ Std Bin | Intervals to right of source sample mean □ in terms of sample standard deviation s: (□ to □+1s), (□+1s to □+2s), (□+2s to □+3s), until all observations fall in an interval. Values to the left of the mean fall into intervals formed in a similar fashion except that endpoints are minus the number of sample standard deviations. (□-4s, □-3s), (□-3s to □-2s), (□-2s to □-1s), (□-1s to □), until all observations fall in an interval. | Shannon's Entropy | Positive and negative standard deviation intervals should be considered separately. Grafstrom (2010) consider unequal probability sampling designs and support the need to have designs to have large entropy. |
| PAL_ Std Bin | | Pal & Pal's Entropy | |
| SHN_ OStd Bin | Intervals are constructed the same as SHN_StdBin or PAL_StdBin, but the sample standard deviation is replaced by the sample standard deviation of the combined data from all sources. | Shannon's Entropy | The combined data from all sources is a larger sample and maybe more representative of the population distribution. The same combined sample standard deviation is used in computing each source’s entropy. Motivation similar to pooling to improve estimate. |
| PAL_ OStd Bin | | Pal & Pal's Entropy | |
| SHN_ PStd Bin | Intervals are constructed the same as SHN_StdBin or PAL_StdBin, but the sample standard deviation is replaced by the pooled standard deviation using the standard deviations of all sources. | Shannon's Entropy | Standard deviation is often a surrogate measure of risk. Risk-pooling improves coordination of supply chain entities. The pooled estimates is a reasonable improved estimate of the population standard deviation. |
| PAL_ PStd Bin | | Pal & Pal's Entropy | |
| SHN_ Fxd Bin | Intervals are constructed the same as SHN_StdBin or PAL_StdBin, but the sample standard deviation is replaced by a fixed interval length. For this study, since the population mean was chosen to be 1000, the bin length was selected to be 100, which is 1/10 of the mean. | Shannon's Entropy | The Bins method does not rely on the values of the standard deviation. The size of the bins will impact the entropy computation (Jansssens et al. 2006). |
| PAL_ Fxd Bin | | Pal & Pal's Entropy | |

For further comparison, two non-entropy based methods, the traditional average and the inverse variance weighted average, and a SAS generalized maximum entropy, are included.

Table 3.2.2 Data Fusion Methods Not Using Sheu (2010)'s Methodology

| Method | Description of Method | Estimation Procedure | Rationale for Method |
|---|---|---|---|
| ENT_GenMax | This Generalized Maximum Entropy (GME) procedure is a method that weights the errors of the estimation bewteen two bounds for the error term. SAS (2014) introduced this procedure as an alternative to a regression based analysis with the parameters selected to maximize the entropy. The objective funcion is $H(p,w) = - p^T*LN(p) - w^T*LN(w)$ in which the vector p is a weight for the model's parameters and the vector w is a weight for the error terms. | The weights are obtained from Generalized Entropy objective source. Entropy estimates for the source means are averaged. | No assumption about the distribution of the error or data is needed (Mittelhammer, Cardell, & Marsh, 2013). |
| TRD_EquWgt | The average of all source averages is the defined as the estimate of the population mean. | The weights are equal. | All sources are considered to be equally reliable. Method is computationally easy. Used to predict demand when no patterns in the data (Chopra & Meindl, 2015). |
| TRD_VarWgt | The weighted average of all source averages is defined as the estimate of the population mean. | The weights are the normalized inverse variance of all sources. | The source with smaller variance is considered more reliable than those with larger variances. Using an objective function to obtain weights is supported by Liu & Zhang (2011) and Kull & Wacker (2010) |

The SAS entropy method, considered suitable when outliers exist or when the typical assumptions of parametric estimation are violated, implements a linear estimation procedure based on generalized maximum entropy. All data fusion methods in Tables 3.2.1 and 3.2.2 are included in this study. The rationales for the formulation of each method are provided in these tables. The first column of Tables 3.2.1 and 3.2.2 provide the mnemonic to describe the particular method. In Table 3.2.1, the first three letters represent the type of entropy, either Shannon's or Pal & Pal's. In Table 3.2.2 the mnemonic TRD is used to represent "traditional" method and ENT represents the maximum entropy principle based method.

### 3.4. Simulation Study Procedure to Assess Data Fusion Methods

The measure of performance from true mean population is the mean absolute percentage error (MAPE), a measure of prediction accuracy or of estimation accuracy of population parameter or random variable's value. MAPE is computed as (100)*(1/N)* (Fi – Ai) /Ai, in which Fi represents the predicted value and Ai represents the actual value or known parameter of interest.

A sensitivity analysis to bias is conducted. Bias may occur in data for various reasons: censorship, partial information, changing patterns, systematic error, and human error as well as the use of subjective information. Bias may be innocently introduced because a respondent misinterpreted the question or a source of information provided "contaminated" data. Data from sources are generated according to the distributions in Table 3.3.1. The increase in standard deviation of the sources in this table illustrates that the inclusion of more sources come at the expense of a decrease in reliability of the total sample. However, the advantage of using more sources is that a larger sample size is available and thus more information is used.

Table 3.3.1 Data are simulated from sources having either normal or lognormal distributions, with increasing standard deviations to indicate a reduction in reliability.

| Distribution of Source Inputs | | | | | |
|---|---|---|---|---|---|
| | Normal Distribution | | Lognormal Distribution | | |
| Data Sources | Means | Standard Deviation | Means | Standard Deviation | Reliability |

| | | | | | |
|---|---|---|---|---|---|
| Source 1 | 1000 | 50 | 1000 | 39 | Most Reliable |
| Source 2 | 1000 | 100 | 1000 | 74 | |
| Source 3 | 1000 | 150 | 1000 | 125 | |
| Source 4 | 1000 | 200 | 1000 | 161 | |
| Source 5 | 1000 | 250 | 1000 | 350 | |
| Source 6 | 1000 | 300 | 1000 | 490 | |
| Source 7 | 1000 | 350 | 1000 | 1123 | |
| Source 8 | 1000 | 400 | 1000 | 3963 | Least Reliable |

Table 3.3.2 Biases added to source 1 in simulation study. The zero positive and zero negative under each distribution is only one configuration. Bias is added only to source 1.

| Biases for Source 1 | | | |
|---|---|---|---|
| Normal Distribution | | Lognormal Distribution | |
| Positive | Negative | Positive | Negative |
| 0 | 0 | 0 | 0 |
| 100 | -100 | 111 | -100 |
| 300 | -300 | 400 | -300 |
| 600 | -600 | 1500 | -600 |

This study assesses the sensitivity of the proposed data fusion methods to possible bias. This bias in added to the first source, which should be typically weighted the most since it is the most reliable. Adding the bias to less reliable sources will confound the effect of increased variance and bias. Both negative and positive bias effects are assessed since a method may tend to underestimate or overestimate a parameter of interest.

The simulation study is replicated a 1000 times for each of the following conditions to assess the robustness of the methods:

1. Normally or lognormally distributed inputs for the sources.
2. Eight Biases – four of which are negative and four of which are positive,
3. Two levels of number of sources – four or eight and
4. Two levels of number of inputs to each source level – 5 or 15.

These conditions are listed in Tables 3.3.2 and 3.3.3. The study determines the effect of these conditions on the estimation of the mean of a population.

Table 3.3.3 Simulation study configuration conditions

| Simulation Configurations | | |
|---|---|---|
| Conditions | Number of Levels | Levels |
| Distributions | 2 | Normal or Lognormal |
| Biases | 7 | Three negative biases,<br>three positive biases,<br>and one zero bias |
| Sources | 2 | Either 4 sources or 8 sources |
| Sample inputs | 2 | Samples of either 5 or 15 |
| | | |
| Total Conditions | 56 | |

## 3.5. Monte Carlo Simulation Results of Data Fusion Estimation of Mean Using Normally Distributed Inputs with Possible Bias

The first set of tables, that is, Tables 3.4.1 through 3.4.4, illustrates the estimation accuracy of the 13 data fusion procedures under the assumption that the inputs from the sources are normally distributed. The tables display the Tukey significance groupings by attaching the same letter next to procedures that are not significantly different. The MAPE values are low, mostly between 1 and 4 percent when there is no bias. Sheu (2010)'s procedure (SHN_Sheu) is highlighted in each of tables as a benchmark to compare the other methods. Since SHN_Sheu is the primary EBDF method in the SCM literature, methods that perform equally well or better can be considered more robust, particularly if the method is consistent in its performance.

What are the simulation results under the assumption of normally distributed responses reveal about these 13 data fusion methods? The following results are deduced from these four tables:

1. With no bias present, the inverse variance weighted (TRD_VarWgt) is recommended as the most accurate estimator. The Generalized Maximum Entropy Estimator is the least accurate. Using bins constructed with the overall or pooled standard deviation or even the fixed bin procedures (*_OStdBin, *_PStdBin, *_FxdBin) have an advantage over the benchmark SHN_Sheu method.

Table 3.4.1 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 4 sources with 5 inputs per source. Data follow a normal distribution.

| Normal Dist - Bias: None | | | | | | Normal Dist - Bias: Neg 100 | | | | Normal Dist - Bias: Neg 300 | | | | Normal Dist - Bias: Neg 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | | A | | ENT_GenMax | 2.94 | | A | TRD_VarWgt | 6.55 | | A | TRD_VarWgt | 19.31 | | A | TRD_VarWgt | 38.54 |
| B | | A | | SHN_StdBin | 2.86 | | B | SHN_FxdBin | 4.58 | | B | SHN_FxdBin | 13.19 | | B | SHN_FxdBin | 26.37 |
| B | | A | C | SHN_Sheu | 2.69 | C | B | SHN_PStdBin | 4.26 | | C | SHN_PStdBin | 11.88 | | C | SHN_PStdBin | 23.74 |
| B | D | A | C | PAL_StdBin | 2.66 | C | | ENT_GenMax | 4.19 | | D | SHN_OStdBin | 9.63 | | D | PAL_FxdBin | 18.96 |
| B | D | | C | PAL_Sheu | 2.63 | C | D | SHN_OStdBin | 4.10 | | D | PAL_FxdBin | 9.50 | | D | PAL_PStdBin | 18.13 |
| B | D | | C | TRD_EquWgt | 2.61 | E | D | SHN_StdBin | 3.75 | E | D | PAL_PStdBin | 9.09 | | E | ENT_GenMax | 16.21 |
| | D | E | C | PAL_PStdBin | 2.42 | E | F | PAL_FxdBin | 3.60 | E | F | ENT_GenMax | 8.70 | | E | SHN_OStdBin | 15.98 |
| | D | E | C | PAL_OStdBin | 2.42 | E | F | PAL_PStdBin | 3.53 | G | F | PAL_OStdBin | 8.35 | F | E | PAL_OStdBin | 15.40 |
| | D | E | | PAL_FxdBin | 2.39 | E | F | PAL_OStdBin | 3.51 | G | H | SHN_StdBin | 7.81 | F | | TRD_EquWgt | 15.04 |
| | | E | | SHN_OStdBin | 2.27 | E | F | SHN_Sheu | 3.46 | | H | TRD_EquWgt | 7.60 | F | | SHN_StdBin | 15.02 |
| | | E | | SHN_PStdBin | 2.26 | E | F | PAL_StdBin | 3.41 | | H | PAL_StdBin | 7.57 | F | | PAL_StdBin | 14.97 |
| F | | E | | SHN_FxdBin | 2.15 | E | F | PAL_Sheu | 3.39 | | H | PAL_Sheu | 7.55 | F | | PAL_Sheu | 14.96 |
| F | | | | TRD_VarWgt | 1.89 | | F | TRD_EquWgt | 3.37 | | H | SHN_Sheu | 7.54 | F | | SHN_Sheu | 14.85 |

| Normal Dist - Bias: None | | | | | | Normal Dist - Bias: Pos 100 | | | | Normal Dist - Bias: Pos 300 | | | | Normal Dist - Bias: Pos 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | | A | | ENT_GenMax | 2.94 | | A | TRD_VarWgt | 6.52 | | A | TRD_VarWgt | 19.25 | | A | TRD_VarWgt | 38.46 |
| B | | A | | SHN_StdBin | 2.86 | | B | SHN_FxdBin | 4.63 | | B | SHN_FxdBin | 13.25 | | B | SHN_FxdBin | 26.43 |
| B | | A | C | SHN_Sheu | 2.69 | | C | SHN_PStdBin | 4.24 | | C | SHN_PStdBin | 11.91 | | C | SHN_PStdBin | 23.77 |
| B | D | A | C | PAL_StdBin | 2.66 | D | C | SHN_OStdBin | 4.14 | | D | SHN_OStdBin | 9.48 | | D | PAL_FxdBin | 18.93 |
| B | D | | C | PAL_Sheu | 2.63 | D | E | SHN_StdBin | 3.77 | | D | PAL_FxdBin | 9.46 | | D | PAL_PStdBin | 18.09 |
| B | D | | C | TRD_EquWgt | 2.61 | F | E | PAL_FxdBin | 3.61 | | D | PAL_PStdBin | 9.04 | | E | SHN_OStdBin | 15.90 |
| | D | E | C | PAL_PStdBin | 2.42 | F | E | PAL_PStdBin | 3.53 | | E | PAL_OStdBin | 8.24 | F | E | PAL_OStdBin | 15.31 |
| | D | E | C | PAL_OStdBin | 2.42 | F | E | PAL_OStdBin | 3.51 | F | E | SHN_StdBin | 7.83 | F | E | SHN_StdBin | 15.11 |
| | D | E | | PAL_FxdBin | 2.39 | F | E | SHN_Sheu | 3.43 | F | | PAL_StdBin | 7.54 | F | | PAL_StdBin | 14.98 |
| | | E | | SHN_OStdBin | 2.27 | F | E | PAL_StdBin | 3.41 | F | | SHN_Sheu | 7.51 | F | | TRD_EquWgt | 14.96 |
| | | E | | SHN_PStdBin | 2.26 | F | G | PAL_Sheu | 3.35 | F | | TRD_EquWgt | 7.49 | F | | PAL_Sheu | 14.91 |
| F | | E | | SHN_FxdBin | 2.15 | F | G | TRD_EquWgt | 3.33 | F | | PAL_Sheu | 7.48 | F | | SHN_Sheu | 14.84 |
| F | | | | TRD_VarWgt | 1.89 | | G | ENT_GenMax | 3.00 | | G | ENT_GenMax | 6.46 | | G | ENT_GenMax | 13.87 |

Table 3.4.2 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 4 sources with 15 inputs per source.  Data follow a normal distribution.

| Normal Dist - Bias: None | | | Normal Dist - Bias: Neg 100 | | | Normal Dist - Bias: Neg 300 | | | Normal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | ENT_GenMax | 1.57 | A | TRD_VarWgt | 6.77 | A | TRD_VarWgt | 20.45 | A | TRD_VarWgt | 40.97 |
| B A | SHN_Sheu | 1.46 | B | SHN_FxdBin | 5.35 | B | SHN_FxdBin | 16.21 | B | SHN_FxdBin | 32.51 |
| B | PAL_Sheu | 1.43 | C | SHN_PStdBin | 5.03 | C | SHN_PStdBin | 15.26 | C | SHN_PStdBin | 30.60 |
| B | SHN_StdBin | 1.43 | C | SHN_OStdBin | 4.91 | D | SHN_OStdBin | 12.29 | D | PAL_FxdBin | 20.54 |
| B | TRD_EquWgt | 1.42 | D | PAL_FxdBin | 3.36 | E | PAL_FxdBin | 10.22 | E | PAL_PStdBin | 19.70 |
| B | PAL_StdBin | 1.42 | D | ENT_GenMax | 3.22 | F | PAL_PStdBin | 9.80 | F | SHN_OStdBin | 18.08 |
| C | PAL_OStdBin | 1.23 | D | PAL_PStdBin | 3.22 | G | PAL_OStdBin | 8.90 | G | PAL_OStdBin | 15.86 |
| C | PAL_PStdBin | 1.23 | D | PAL_OStdBin | 3.18 | H | ENT_GenMax | 8.17 | G | ENT_GenMax | 15.69 |
| C | PAL_FxdBin | 1.22 | E | SHN_Sheu | 2.56 | I | TRD_EquWgt | 7.39 | H | TRD_EquWgt | 14.89 |
| D | SHN_OStdBin | 0.98 | E | SHN_StdBin | 2.53 | I | PAL_StdBin | 7.37 | H | PAL_StdBin | 14.85 |
| D | SHN_FxdBin | 0.98 | E | PAL_Sheu | 2.53 | I | PAL_Sheu | 7.36 | H | PAL_Sheu | 14.84 |
| D | SHN_PStdBin | 0.98 | E | TRD_EquWgt | 2.52 | I | SHN_StdBin | 7.36 | H | SHN_StdBin | 14.83 |
| D | TRD_VarWgt | 0.91 | E | PAL_StdBin | 2.52 | I | SHN_Sheu | 7.33 | H | SHN_Sheu | 14.78 |

| Normal Dist - Bias: None | | | Normal Dist - Bias: Pos 100 | | | Normal Dist - Bias: Pos 300 | | | Normal Dist - Bias: Pos 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | ENT_GenMax | 1.57 | A | TRD_VarWgt | 6.90 | A | TRD_VarWgt | 20.58 | A | TRD_VarWgt | 41.10 |
| B A | SHN_Sheu | 1.46 | B | SHN_FxdBin | 5.51 | B | SHN_FxdBin | 16.37 | B | SHN_FxdBin | 32.66 |
| B | PAL_Sheu | 1.43 | C | SHN_PStdBin | 5.19 | C | SHN_PStdBin | 15.42 | C | SHN_PStdBin | 30.75 |
| B | SHN_StdBin | 1.43 | C | SHN_OStdBin | 5.05 | D | SHN_OStdBin | 12.44 | D | PAL_FxdBin | 20.74 |
| B | TRD_EquWgt | 1.42 | D | PAL_FxdBin | 3.55 | E | PAL_FxdBin | 10.42 | E | PAL_PStdBin | 19.90 |
| B | PAL_StdBin | 1.42 | D | PAL_PStdBin | 3.41 | F | PAL_PStdBin | 10.00 | F | SHN_OStdBin | 18.25 |
| C | PAL_OStdBin | 1.23 | D | PAL_OStdBin | 3.37 | G | PAL_OStdBin | 9.09 | G | PAL_OStdBin | 16.06 |
| C | PAL_PStdBin | 1.23 | E | SHN_Sheu | 2.77 | H | TRD_EquWgt | 7.61 | H | TRD_EquWgt | 15.11 |
| C | PAL_FxdBin | 1.22 | E | SHN_StdBin | 2.73 | H | PAL_StdBin | 7.59 | H | PAL_StdBin | 15.08 |
| D | SHN_OStdBin | 0.98 | E | PAL_Sheu | 2.72 | H | PAL_Sheu | 7.59 | H | PAL_Sheu | 15.07 |
| D | SHN_FxdBin | 0.98 | E | TRD_EquWgt | 2.72 | H | SHN_Sheu | 7.59 | H | SHN_StdBin | 15.06 |
| D | SHN_PStdBin | 0.98 | E | PAL_StdBin | 2.71 | H | SHN_StdBin | 7.59 | H | SHN_Sheu | 15.05 |
| D | TRD_VarWgt | 0.91 | F | ENT_GenMax | 2.16 | I | ENT_GenMax | 6.85 | I | ENT_GenMax | 14.36 |

Table 3.4.3 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 8 sources with 5 inputs per source. Data follow a normal distribution.

| Normal Dist - Bias: None | | | Normal Dist - Bias: Neg 100 | | | Normal Dist - Bias: Neg 300 | | | Normal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | ENT_GenMax | 4.20 | A | TRD_VarWgt | 6.00 | A | TRD_VarWgt | 17.60 | A | TRD_VarWgt | 35.11 |
| A | SHN_StdBin | 3.93 | B | ENT_GenMax | 4.87 | B | SHN_FxdBin | 8.91 | B | SHN_FxdBin | 17.87 |
| B | SHN_Sheu | 3.54 | C | SHN_StdBin | 4.07 | C | ENT_GenMax | 6.76 | C | SHN_PStdBin | 11.34 |
| B | PAL_StdBin | 3.46 | D | SHN_Sheu | 3.62 | D | SHN_PStdBin | 5.90 | D | PAL_FxdBin | 10.50 |
| B | PAL_Sheu | 3.42 | E D | PAL_StdBin | 3.54 | E D | SHN_OStdBin | 5.59 | E D | ENT_GenMax | 10.32 |
| C B | TRD_EquWgt | 3.37 | E D | PAL_Sheu | 3.50 | E D | PAL_FxdBin | 5.42 | E | SHN_OStdBin | 9.70 |
| C D | PAL_PStdBin | 3.06 | E D | SHN_FxdBin | 3.50 | E F | SHN_StdBin | 5.20 | F | PAL_PStdBin | 8.81 |
| C D | PAL_OStdBin | 3.05 | E D | TRD_EquWgt | 3.45 | G F | PAL_PStdBin | 4.80 | F | PAL_OStdBin | 8.27 |
| D | PAL_FxdBin | 3.01 | E D | PAL_FxdBin | 3.25 | G F | PAL_OStdBin | 4.72 | G F | SHN_StdBin | 8.03 |
| E D | SHN_OStdBin | 2.83 | E D | SHN_PStdBin | 3.23 | G | SHN_Sheu | 4.59 | G | SHN_Sheu | 7.47 |
| E D | SHN_PStdBin | 2.81 | E | PAL_PStdBin | 3.23 | G | PAL_StdBin | 4.51 | G | PAL_StdBin | 7.44 |
| E | SHN_FxdBin | 2.49 | E | PAL_OStdBin | 3.22 | G | PAL_Sheu | 4.45 | G | PAL_Sheu | 7.41 |
| F | TRD_VarWgt | 1.88 | E | SHN_OStdBin | 3.21 | G | TRD_EquWgt | 4.43 | G | TRD_EquWgt | 7.41 |

| Normal Dist - Bias: None | | | Normal Dist - Bias: Pos 100 | | | Normal Dist - Bias: Pos 300 | | | Normal Dist - Bias: Pos 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | ENT_GenMax | 4.20 | A | TRD_VarWgt | 5.99 | A | TRD_VarWgt | 17.53 | A | TRD_VarWgt | 35.03 |
| A | SHN_StdBin | 3.93 | B | SHN_StdBin | 4.21 | B | SHN_FxdBin | 9.26 | B | SHN_FxdBin | 18.22 |
| B | SHN_Sheu | 3.54 | C | SHN_FxdBin | 3.77 | C | SHN_PStdBin | 6.01 | C | SHN_PStdBin | 11.51 |
| B | PAL_StdBin | 3.46 | D C | SHN_Sheu | 3.75 | C | SHN_OStdBin | 5.83 | C | PAL_FxdBin | 10.91 |
| B | PAL_Sheu | 3.42 | D C E | ENT_GenMax | 3.74 | C | PAL_FxdBin | 5.79 | D | SHN_OStdBin | 9.93 |
| C B | TRD_EquWgt | 3.37 | D F C E | PAL_StdBin | 3.69 | D C | SHN_StdBin | 5.57 | E D | PAL_PStdBin | 9.15 |
| C D | PAL_PStdBin | 3.06 | G D F C E | PAL_Sheu | 3.63 | D E | PAL_PStdBin | 5.04 | E F | PAL_OStdBin | 8.62 |
| C D | PAL_OStdBin | 3.05 | G D F C E | TRD_EquWgt | 3.59 | E | PAL_OStdBin | 5.01 | E F G | SHN_StdBin | 8.51 |
| D | PAL_FxdBin | 3.01 | G D F C E | PAL_FxdBin | 3.44 | E | SHN_Sheu | 4.92 | F G | PAL_StdBin | 7.91 |
| E D | SHN_OStdBin | 2.83 | G D F E | PAL_OStdBin | 3.36 | E | PAL_StdBin | 4.88 | F G | SHN_Sheu | 7.86 |
| E D | SHN_PStdBin | 2.81 | G F E | PAL_PStdBin | 3.35 | E | PAL_Sheu | 4.79 | G | TRD_EquWgt | 7.80 |
| E | SHN_FxdBin | 2.49 | G F | SHN_OStdBin | 3.31 | E | TRD_EquWgt | 4.76 | G | PAL_Sheu | 7.80 |
| F | TRD_VarWgt | 1.88 | G | SHN_PStdBin | 3.29 | F | ENT_GenMax | 3.64 | H | ENT_GenMax | 5.40 |

Table 3.4.4 Tukey grouping using Mean Absolute Percent Error in estimation of population mean employing 8 sources with 15 inputs per source. Data follow a normal distribution.

| Normal Dist - Bias: None | | | Normal Dist - Bias: Neg 100 | | | Normal Dist - Bias: Neg 300 | | | Normal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | ENT_GenMax | 2.63 | A | TRD_VarWgt | 6.29 | A | TRD_VarWgt | 18.99 | A | TRD_VarWgt | 18.99 |
| B | SHN_Sheu | 1.88 | B | SHN_FxdBin | 4.01 | B | SHN_FxdBin | 12.24 | B | SHN_FxdBin | 12.24 |
| B | SHN_StdBin | 1.83 | C | ENT_GenMax | 3.54 | C | SHN_PStdBin | 7.58 | C | SHN_PStdBin | 7.58 |
| B | PAL_Sheu | 1.83 | D | SHN_PStdBin | 2.52 | D | SHN_OStdBin | 7.15 | D | SHN_OStdBin | 7.15 |
| B | TRD_EquWgt | 1.82 | D | SHN_OStdBin | 2.50 | E | ENT_GenMax | 5.89 | E | ENT_GenMax | 5.89 |
| B | PAL_StdBin | 1.82 | E | PAL_FxdBin | 2.20 | E | PAL_FxdBin | 5.88 | E | PAL_FxdBin | 5.88 |
| C | PAL_FxdBin | 1.55 | E | SHN_Sheu | 2.09 | F | PAL_PStdBin | 4.81 | F | PAL_PStdBin | 4.81 |
| C | PAL_OStdBin | 1.54 | E | SHN_StdBin | 2.04 | F | PAL_OStdBin | 4.70 | F | PAL_OStdBin | 4.70 |
| C | PAL_PStdBin | 1.54 | E | PAL_Sheu | 2.02 | G | SHN_Sheu | 3.71 | G | SHN_Sheu | 3.71 |
| D | SHN_OStdBin | 1.18 | E | PAL_StdBin | 2.02 | G | SHN_StdBin | 3.68 | G | SHN_StdBin | 3.68 |
| D | SHN_PStdBin | 1.18 | E | TRD_EquWgt | 2.02 | G | TRD_EquWgt | 3.67 | G | TRD_EquWgt | 3.67 |
| D | SHN_FxdBin | 1.09 | E | PAL_OStdBin | 1.97 | G | PAL_StdBin | 3.67 | G | PAL_StdBin | 3.67 |
| E | TRD_VarWgt | 0.88 | E | PAL_PStdBin | 1.97 | G | PAL_Sheu | 3.66 | G | PAL_Sheu | 3.66 |

| Normal Dist - Bias: None | | | Normal Dist - Bias: Pos 100 | | | Normal Dist - Bias: Pos 300 | | | Normal Dist - Bias: Pos 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | ENT_GenMax | 2.63 | A | TRD_VarWgt | 6.42 | A | TRD_VarWgt | 19.13 | A | TRD_VarWgt | 38.19 |
| B | SHN_Sheu | 1.88 | B | SHN_FxdBin | 4.25 | B | SHN_FxdBin | 12.48 | B | SHN_FxdBin | 24.84 |
| B | SHN_StdBin | 1.83 | C | SHN_PStdBin | 2.71 | C | SHN_PStdBin | 7.79 | C | SHN_PStdBin | 15.48 |
| B | PAL_Sheu | 1.83 | C | SHN_OStdBin | 2.70 | D | SHN_OStdBin | 7.35 | D | SHN_OStdBin | 12.91 |
| B | TRD_EquWgt | 1.82 | D | PAL_FxdBin | 2.44 | E | PAL_FxdBin | 6.17 | E | PAL_FxdBin | 12.19 |
| B | PAL_StdBin | 1.82 | E D | SHN_Sheu | 2.25 | F | PAL_PStdBin | 5.09 | F | PAL_PStdBin | 10.03 |
| C | PAL_FxdBin | 1.55 | E | SHN_StdBin | 2.20 | F | PAL_OStdBin | 4.98 | G | PAL_OStdBin | 9.32 |
| C | PAL_OStdBin | 1.54 | E | PAL_Sheu | 2.19 | G | SHN_Sheu | 4.05 | H | TRD_EquWgt | 7.66 |
| C | PAL_PStdBin | 1.54 | E | PAL_StdBin | 2.18 | G | SHN_StdBin | 4.00 | H | PAL_Sheu | 7.65 |
| D | SHN_OStdBin | 1.18 | E | PAL_PStdBin | 2.18 | G | PAL_Sheu | 4.00 | H | PAL_StdBin | 7.64 |
| D | SHN_PStdBin | 1.18 | E | PAL_OStdBin | 2.17 | G | TRD_EquWgt | 4.00 | H | SHN_StdBin | 7.64 |
| D | SHN_FxdBin | 1.09 | E | TRD_EquWgt | 2.17 | G | PAL_StdBin | 3.99 | H | SHN_Sheu | 7.64 |
| E | TRD_VarWgt | 0.88 | E | ENT_GenMax | 2.06 | H | ENT_GenMax | 2.37 | I | ENT_GenMax | 5.40 |

2. As the bias becomes more pronounced, either negative or positive, the SHN_Sheu method's performance increases on a relative basis. Clearly, the MAPE increases, in general, as the absolute value of the bias increases. The TRD_VarWgt's performance deteriorates as the absolute value of the bias increases. The use of Pal & Pal entropy does not appear to be significantly different from the use of Shannon's entropy, particularly for the SHN_Sheu. Many of the procedures compared only on the basis of Shannon or Pal & Pal entropies are not significantly different in performance although the use of Pal & Pal entropy generally improves performance relatively.

3. The Generalized Maximum Entropy method (ENT_GenMax) is the best performer in the presence of large bias. The equal weight procedure (TRD_EquWgt) generally is among the better performers as bias is introduced.

4. Sensitivity analysis, conducted by increasing the number of inputs per data source, shows that the relative performance of the methods is not altered substantially.

## 3.6. Simulation Results of Data Fusion Estimation of Mean Parameter Using Lognormally Distributed Inputs with Possible Bias

What are the simulation results under the assumption of lognormally distributed responses reveal about these 13 data fusion methods? The following results are deduced from Tables 3.5.1 through 3.5.4 which demonstrate the robustness of the data fusion methods, ranked by MAPE. The best methods are at the bottom and the worst are at the top of the tables. The results should be interpreted in conjunction with the Tukey grouping.

Table 3.5.1 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 4 sources with 5 inputs from a lognormal distribution

**Lognormal Dist - Bias: None**

| Pair-wise Signif | | Method | MAPE (%) |
|---|---|---|---|
| | A | ENT_GenMax | 2.32 |
| | A | SHN_StdBin | 2.30 |
| | A | SHN_Sheu | 2.16 |
| B | A | PAL_StdBin | 2.14 |
| B | A | PAL_Sheu | 2.11 |
| B | A | TRD_EquWgt | 2.09 |
| B | C | PAL_OStdBin | 1.91 |
| B | C | PAL_PStdBin | 1.91 |
| B | C | PAL_FxdBin | 1.91 |
| | C | SHN_OStdBin | 1.76 |
| | C | SHN_FxdBin | 1.75 |
| | C | SHN_PStdBin | 1.75 |
| | D | TRD_VarWgt | 1.49 |

**Lognormal Dist - Bias: Neg 100**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | A | | TRD_VarWgt | 7.01 |
| | B | | SHN_FxdBin | 4.24 |
| | B | | SHN_PStdBin | 4.20 |
| C | B | | SHN_OStdBin | 4.01 |
| C | | | ENT_GenMax | 3.78 |
| | D | | SHN_StdBin | 3.37 |
| E | D | | PAL_FxdBin | 3.35 |
| E | D | | PAL_PStdBin | 3.35 |
| E | D | F | PAL_OStdBin | 3.30 |
| E | D | F | SHN_Sheu | 3.12 |
| E | D | F | PAL_StdBin | 3.07 |
| E | | F | PAL_Sheu | 3.05 |
| | | F | TRD_EquWgt | 3.03 |

**Lognormal Dist - Bias: Neg 300**

| Pair-wise Signif | | Method | MAPE (%) |
|---|---|---|---|
| A | | TRD_VarWgt | 23.06 |
| B | | SHN_FxdBin | 12.15 |
| B | | SHN_PStdBin | 12.04 |
| C | | PAL_FxdBin | 9.19 |
| C | | SHN_OStdBin | 9.17 |
| C | | PAL_PStdBin | 9.15 |
| D | | ENT_GenMax | 8.51 |
| D | | PAL_OStdBin | 8.17 |
| E | | SHN_StdBin | 7.61 |
| E | | TRD_EquWgt | 7.54 |
| E | | PAL_Sheu | 7.50 |
| E | | PAL_StdBin | 7.49 |
| E | | SHN_Sheu | 7.46 |

**Lognormal Dist - Bias: Neg 600**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | A | | TRD_VarWgt | 53.84 |
| | B | | SHN_FxdBin | 24.14 |
| | B | | SHN_PStdBin | 24.04 |
| | C | | PAL_FxdBin | 18.28 |
| | C | | PAL_PStdBin | 18.25 |
| | D | | ENT_GenMax | 15.99 |
| E | D | | SHN_OStdBin | 15.43 |
| E | | | PAL_OStdBin | 15.18 |
| E | | | TRD_EquWgt | 15.03 |
| E | | | PAL_Sheu | 14.95 |
| E | | | PAL_StdBin | 14.94 |
| E | | | SHN_StdBin | 14.88 |
| E | | | SHN_Sheu | 14.85 |

**Lognormal Dist - Bias: None**

| Pair-wise Signif | | Method | MAPE (%) |
|---|---|---|---|
| | A | ENT_GenMax | 2.32 |
| | A | SHN_StdBin | 2.30 |
| | A | SHN_Sheu | 2.16 |
| B | A | PAL_StdBin | 2.14 |
| B | A | PAL_Sheu | 2.11 |
| B | A | TRD_EquWgt | 2.09 |
| B | C | PAL_OStdBin | 1.91 |
| B | C | PAL_PStdBin | 1.91 |
| B | C | PAL_FxdBin | 1.91 |
| | C | SHN_OStdBin | 1.76 |
| | C | SHN_FxdBin | 1.75 |
| | C | SHN_PStdBin | 1.75 |
| | D | TRD_VarWgt | 1.49 |

**Lognormal Dist - Bias: Pos 111**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | A | | TRD_VarWgt | 6.49 |
| | B | | SHN_FxdBin | 4.35 |
| | B | | SHN_PStdBin | 4.27 |
| | B | | SHN_OStdBin | 4.03 |
| | C | | SHN_StdBin | 3.57 |
| D | C | | PAL_FxdBin | 3.48 |
| D | C | E | PAL_PStdBin | 3.44 |
| D | C | E | PAL_OStdBin | 3.40 |
| D | C | E | SHN_Sheu | 3.26 |
| D | C | E | PAL_StdBin | 3.23 |
| D | | E | PAL_Sheu | 3.17 |
| | | E | TRD_EquWgt | 3.14 |
| | F | | ENT_GenMax | 2.59 |

**Lognormal Dist - Bias: Pos 428**

| Pair-wise Signif | | Method | MAPE (%) |
|---|---|---|---|
| | A | TRD_VarWgt | 20.88 |
| | B | SHN_FxdBin | 15.94 |
| | B | SHN_PStdBin | 15.51 |
| | C | PAL_FxdBin | 12.60 |
| | C | PAL_PStdBin | 12.46 |
| | D | SHN_OStdBin | 11.52 |
| E | D | PAL_OStdBin | 11.00 |
| E | | SHN_StdBin | 10.81 |
| E | | PAL_StdBin | 10.71 |
| E | | TRD_EquWgt | 10.69 |
| E | | PAL_Sheu | 10.67 |
| E | | SHN_Sheu | 10.64 |
| | F | ENT_GenMax | 9.70 |

**Lognormal Dist - Bias: Pos 1500**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | A | | TRD_VarWgt | 41.36 |
| B | A | | SHN_PStdBin | 41.02 |
| B | A | C | SHN_FxdBin | 40.08 |
| B | D | C | PAL_PStdBin | 39.08 |
| | D | C | PAL_FxdBin | 38.70 |
| E | D | | TRD_EquWgt | 37.48 |
| E | D | | PAL_OStdBin | 37.47 |
| E | D | | SHN_OStdBin | 37.46 |
| E | D | | PAL_StdBin | 37.43 |
| E | D | | SHN_StdBin | 37.41 |
| E | D | | PAL_Sheu | 37.35 |
| E | D | | SHN_Sheu | 37.19 |
| E | | | ENT_GenMax | 36.46 |

Table 3.5.2 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 4 sources with 5 inputs from a lognormal distribution

| Lognormal Dist - Bias: None | | | | Lognormal Dist - Bias: Neg 100 | | | Lognormal Dis - Bias: Neg 300 | | | Lognormal Dist - Bias: Neg 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | A | ENT_GenMax | 1.29 | A | TRD_VarWgt | 7.24 | A | TRD_VarWgt | 24.27 | | A | TRD_VarWgt | 55.59 |
| B | A | SHN_Sheu | 1.24 | B | SHN_FxdBin | 5.28 | B | SHN_FxdBin | 16.01 | | B | SHN_FxdBin | 32.03 |
| B | | SHN_StdBin | 1.15 | C | SHN_PStdBin | 5.06 | C | SHN_PStdBin | 15.37 | | C | SHN_PStdBin | 30.87 |
| B | | PAL_Sheu | 1.15 | D | SHN_OStdBin | 4.83 | D | SHN_OStdBin | 11.15 | | D | PAL_FxdBin | 20.14 |
| B | | PAL_StdBin | 1.14 | E | PAL_FxdBin | 3.30 | E | PAL_FxdBin | 10.04 | | D | PAL_PStdBin | 19.80 |
| B | | TRD_EquWgt | 1.14 | E | PAL_PStdBin | 3.24 | E | PAL_PStdBin | 9.86 | | E | SHN_OStdBin | 16.68 |
| | C | PAL_OStdBin | 0.97 | E | ENT_GenMax | 3.22 | F | PAL_OStdBin | 8.60 | | F | ENT_GenMax | 15.70 |
| | C | PAL_PStdBin | 0.97 | E | PAL_OStdBin | 3.17 | G | ENT_GenMax | 8.21 | | F | PAL_OStdBin | 15.41 |
| | C | PAL_FxdBin | 0.97 | F | SHN_Sheu | 2.52 | H | TRD_EquWgt | 7.41 | | G | TRD_EquWgt | 14.92 |
| | D | SHN_OStdBin | 0.77 | F | TRD_EquWgt | 2.46 | H | PAL_StdBin | 7.36 | H | G | PAL_StdBin | 14.81 |
| | D | SHN_PStdBin | 0.77 | F | PAL_Sheu | 2.46 | H | PAL_Sheu | 7.36 | H | G | PAL_Sheu | 14.81 |
| | D | SHN_FxdBin | 0.76 | F | SHN_StdBin | 2.45 | H | SHN_StdBin | 7.32 | H | G | SHN_StdBin | 14.74 |
| | D | TRD_VarWgt | 0.71 | F | PAL_StdBin | 2.45 | H | SHN_Sheu | 7.25 | H | | SHN_Sheu | 14.53 |

| Lognormal Dist - Bias: None | | | | Lognormal Dist - Bias: Pos 111 | | | Lognormal Dist - Bias: Pos 428 | | | Lognormal Dist - Bias: Pos 1500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | A | ENT_GenMax | 1.29 | A | TRD_VarWgt | 7.04 | A | TRD_VarWgt | 22.23 | | A | TRD_VarWgt | 40.76 |
| B | A | SHN_Sheu | 1.24 | B | SHN_FxdBin | 5.62 | B | SHN_FxdBin | 18.76 | B | A | SHN_PStdBin | 39.80 |
| B | | SHN_StdBin | 1.15 | B | SHN_PStdBin | 5.42 | B | SHN_PStdBin | 18.72 | B | C | SHN_FxdBin | 39.23 |
| B | | PAL_Sheu | 1.15 | C | SHN_OStdBin | 5.19 | C | SHN_OStdBin | 13.63 | B | C | PAL_PStdBin | 39.06 |
| B | | PAL_StdBin | 1.14 | D | PAL_FxdBin | 3.72 | C | PAL_FxdBin | 13.49 | D | C | PAL_FxdBin | 38.64 |
| B | | TRD_EquWgt | 1.14 | D | PAL_PStdBin | 3.66 | C | PAL_PStdBin | 13.42 | D | E | TRD_EquWgt | 37.60 |
| | C | PAL_OStdBin | 0.97 | D | PAL_OStdBin | 3.58 | D | PAL_OStdBin | 11.65 | D | E | PAL_OStdBin | 37.56 |
| | C | PAL_PStdBin | 0.97 | E | SHN_Sheu | 2.93 | E | TRD_EquWgt | 10.81 | D | E | SHN_OStdBin | 37.53 |
| | C | PAL_FxdBin | 0.97 | E | TRD_EquWgt | 2.89 | E | PAL_StdBin | 10.74 | | E | PAL_StdBin | 37.36 |
| | D | SHN_OStdBin | 0.77 | E | PAL_StdBin | 2.87 | E | PAL_Sheu | 10.73 | | E | PAL_Sheu | 37.34 |
| | D | SHN_PStdBin | 0.77 | E | PAL_Sheu | 2.87 | E | SHN_StdBin | 10.69 | | E | SHN_StdBin | 37.18 |
| | D | SHN_FxdBin | 0.76 | E | SHN_StdBin | 2.87 | E | SHN_Sheu | 10.56 | | E | ENT_GenMax | 36.77 |
| | D | TRD_VarWgt | 0.71 | F | ENT_GenMax | 2.16 | F | ENT_GenMax | 10.00 | | E | SHN_Sheu | 36.57 |

Table 3.5.3 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 8 sources with 5 inputs from a lognormal distribution

| Lognormal Dist - Bias: None | | | Lognormal Dist - Bias: Neg 100 | | | Lognormal Dist - Bias: Neg 300 | | | Lognormal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | SHN_StdBin | 23.90 | A | SHN_StdBin | 24.24 | A | SHN_StdBin | 25.09 | A | TRD_VarWgt | 53.65 |
| B | PAL_StdBin | 15.99 | B | PAL_StdBin | 16.35 | A | TRD_VarWgt | 23.35 | B | SHN_StdBin | 26.66 |
| C B | SHN_Sheu | 14.85 | C B | SHN_Sheu | 15.20 | B | PAL_StdBin | 17.26 | C B | SHN_FxdBin | 23.06 |
| C B D | ENT_GenMax | 13.79 | C B | ENT_GenMax | 14.86 | B | ENT_GenMax | 17.05 | C D | ENT_GenMax | 20.45 |
| C B D | PAL_Sheu | 13.48 | C B | PAL_Sheu | 13.83 | B | SHN_Sheu | 16.08 | E D | PAL_StdBin | 18.97 |
| C B D | SHN_OStdBin | 12.67 | C B | SHN_OStdBin | 13.42 | B | SHN_FxdBin | 15.63 | E D | SHN_PStdBin | 18.61 |
| C B D | TRD_EquWgt | 12.42 | C B | SHN_PStdBin | 13.23 | B | SHN_OStdBin | 15.27 | E D | SHN_OStdBin | 18.51 |
| C B D | SHN_PStdBin | 12.39 | C B | PAL_OStdBin | 12.84 | B | SHN_PStdBin | 15.19 | E D | SHN_Sheu | 17.87 |
| C B D | PAL_OStdBin | 12.33 | C B | PAL_PStdBin | 12.78 | B | PAL_Sheu | 14.76 | E D | PAL_FxdBin | 17.53 |
| C B D | PAL_PStdBin | 12.24 | C B | TRD_EquWgt | 12.78 | B | PAL_OStdBin | 14.20 | E D | PAL_PStdBin | 16.82 |
| C D | PAL_FxdBin | 11.54 | C | PAL_FxdBin | 12.15 | B | PAL_PStdBin | 14.19 | E D | PAL_OStdBin | 16.77 |
| D | SHN_FxdBin | 10.23 | C D | SHN_FxdBin | 11.56 | B | PAL_FxdBin | 13.98 | E D | PAL_Sheu | 16.67 |
| E | TRD_VarWgt | 2.33 | D | TRD_VarWgt | 7.95 | B | TRD_EquWgt | 13.77 | E | TRD_EquWgt | 15.78 |

| Lognormal Dist - Bias: None | | | Lognormal Dist - Bias: Pos 111 | | | Lognormal Dist - Bias: Pos 428 | | | Lognormal Dist - Bias: Pos 1500 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| A | SHN_StdBin | 23.90 | A | SHN_StdBin | 23.67 | A | SHN_StdBin | 23.85 | A | TRD_VarWgt | 35.85 |
| B | PAL_StdBin | 15.99 | B | PAL_StdBin | 15.69 | B | TRD_VarWgt | 18.42 | B | SHN_FxdBin | 30.37 |
| C B | SHN_Sheu | 14.85 | C B | SHN_Sheu | 14.56 | C B | PAL_StdBin | 15.48 | B | SHN_StdBin | 28.75 |
| C B D | ENT_GenMax | 13.79 | C B D | PAL_Sheu | 13.18 | C B | SHN_FxdBin | 14.70 | C | PAL_FxdBin | 23.32 |
| C B D | PAL_Sheu | 13.48 | C B D | ENT_GenMax | 12.67 | C B | SHN_Sheu | 14.39 | C | SHN_PStdBin | 22.27 |
| C B D | SHN_OStdBin | 12.67 | C B D | TRD_EquWgt | 12.12 | C D | PAL_Sheu | 12.97 | C | PAL_StdBin | 21.67 |
| C B D | TRD_EquWgt | 12.42 | C B D | SHN_OStdBin | 12.01 | C D | TRD_EquWgt | 11.94 | C | SHN_Sheu | 20.68 |
| C B D | SHN_PStdBin | 12.39 | C B D | PAL_OStdBin | 11.87 | C D | PAL_FxdBin | 11.80 | C | PAL_PStdBin | 20.63 |
| C B D | PAL_OStdBin | 12.33 | C B D | PAL_PStdBin | 11.76 | C D | SHN_OStdBin | 11.74 | C | PAL_Sheu | 20.02 |
| C B D | PAL_PStdBin | 12.24 | C B D | SHN_PStdBin | 11.66 | C D | PAL_OStdBin | 11.59 | C | TRD_EquWgt | 19.89 |
| C D | PAL_FxdBin | 11.54 | C D | PAL_FxdBin | 11.14 | C D | PAL_PStdBin | 11.37 | C | PAL_OStdBin | 19.47 |
| D | SHN_FxdBin | 10.23 | D | SHN_FxdBin | 9.91 | C D | SHN_PStdBin | 11.34 | C | SHN_OStdBin | 19.26 |
| E | TRD_VarWgt | 2.33 | E | TRD_VarWgt | 5.71 | D | ENT_GenMax | 9.75 | D | ENT_GenMax | 8.43 |

Table 3.5.4 Tukey grouping using Mean Absolute Percent Error in estimation of population mean for 8 sources with 15 inputs from a lognormal distribution

| Lognormal Dist - Bias: | | | | Lognormal Dist - Bias: Neg 100 | | | | | Lognormal Dist - Bias: Neg 300 | | | | | Lognormal Dist - Bias: Neg 600 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) |
| | A | SHN_Sheu | 30.90 | | A | | SHN_Sheu | 31.07 | | A | | SHN_Sheu | 31.45 | | A | | TRD_VarWgt | 55.39 |
| | B | SHN_StdBin | 19.17 | | B | | SHN_StdBin | 19.46 | | B | | TRD_VarWgt | 24.12 | | B | | SHN_Sheu | 32.10 |
| | C | ENT_GenMax | 14.29 | | C | | ENT_GenMax | 15.52 | | C | | SHN_StdBin | 20.15 | | C | | SHN_FxdBin | 26.50 |
| | D | PAL_StdBin | 11.26 | | D | | PAL_StdBin | 11.58 | | C | | ENT_GenMax | 18.00 | | D | | ENT_GenMax | 21.72 |
| E | D | PAL_Sheu | 10.08 | E | D | | PAL_Sheu | 10.38 | | D | | SHN_FxdBin | 13.68 | | D | | SHN_StdBin | 21.38 |
| E | F | SHN_OStdBin | 8.07 | E | D | F | SHN_OStdBin | 8.95 | E | D | | PAL_StdBin | 12.41 | | E | | SHN_PStdBin | 15.37 |
| E | F | SHN_PStdBin | 8.02 | E | D | F | SHN_PStdBin | 8.93 | E | D | F | SHN_PStdBin | 11.29 | | E | | SHN_OStdBin | 15.31 |
| E | F | PAL_OStdBin | 7.73 | E | G | F | PAL_OStdBin | 8.28 | E | D | F | SHN_OStdBin | 11.29 | | E | | PAL_StdBin | 14.06 |
| E | F | PAL_PStdBin | 7.70 | E | G | F | PAL_PStdBin | 8.26 | E | D | F | PAL_Sheu | 11.21 | F | E | | PAL_FxdBin | 14.03 |
| E | F | TRD_EquWgt | 7.48 | E | G | F | TRD_EquWgt | 7.84 | E | | F | PAL_OStdBin | 9.76 | F | E | | PAL_Sheu | 12.95 |
| G | F | PAL_FxdBin | 6.11 | | G | F | TRD_VarWgt | 7.31 | E | | F | PAL_PStdBin | 9.76 | F | E | | PAL_PStdBin | 12.58 |
| G | H | SHN_FxdBin | 3.26 | | G | F | PAL_FxdBin | 6.84 | | | F | PAL_FxdBin | 9.25 | F | E | | PAL_OStdBin | 12.57 |
| | H | TRD_VarWgt | 0.76 | | G | | SHN_FxdBin | 5.79 | | | F | TRD_EquWgt | 8.93 | F | | | TRD_EquWgt | 11.18 |

| Lognormal Dist - Bias: | | | | Lognormal Dist - Bias: Pos 111 | | | | | Lognormal Dist - Bias: Pos 428 | | | | | Lognormal Dist - Bias: Pos 1500 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) |
| | A | SHN_Sheu | 30.90 | | A | | SHN_Sheu | 30.75 | | A | | SHN_Sheu | 30.52 | | A | | TRD_VarWgt | 38.46 |
| | B | SHN_StdBin | 19.17 | | B | | SHN_StdBin | 18.89 | | B | | TRD_VarWgt | 21.28 | | B | | SHN_Sheu | 31.34 |
| | C | ENT_GenMax | 14.29 | | C | | ENT_GenMax | 12.91 | | B | | SHN_StdBin | 18.49 | | C | | SHN_FxdBin | 27.95 |
| | D | PAL_StdBin | 11.26 | D | C | D | PAL_StdBin | 10.99 | | C | | SHN_FxdBin | 14.19 | | C | | SHN_PStdBin | 26.02 |
| E | D | PAL_Sheu | 10.08 | D | E | D | PAL_Sheu | 9.85 | | D | | PAL_StdBin | 10.97 | | D | | SHN_OStdBin | 22.49 |
| E | F | SHN_OStdBin | 8.07 | F | E | D | SHN_OStdBin | 7.57 | E | D | | PAL_Sheu | 9.98 | | D | | PAL_FxdBin | 22.16 |
| E | F | SHN_PStdBin | 8.02 | F | E | D | SHN_PStdBin | 7.46 | E | D | | SHN_PStdBin | 9.58 | E | D | | SHN_StdBin | 21.72 |
| E | F | PAL_OStdBin | 7.73 | F | E | D | PAL_OStdBin | 7.35 | E | D | | SHN_OStdBin | 9.42 | E | D | F | PAL_PStdBin | 20.76 |
| E | F | PAL_PStdBin | 7.70 | F | E | D | PAL_PStdBin | 7.29 | E | D | | ENT_GenMax | 9.06 | E | D | F | PAL_OStdBin | 19.31 |
| E | F | TRD_EquWgt | 7.48 | F | E | D | TRD_EquWgt | 7.25 | E | D | | PAL_FxdBin | 8.48 | E | | F | TRD_EquWgt | 18.86 |
| G | F | PAL_FxdBin | 6.11 | F | | D | TRD_VarWgt | 6.65 | E | D | | PAL_OStdBin | 7.94 | | | F | PAL_StdBin | 18.38 |
| G | H | SHN_FxdBin | 3.26 | F | | D | PAL_FxdBin | 5.89 | E | D | | PAL_PStdBin | 7.88 | | | F | PAL_Sheu | 17.87 |
| | H | TRD_VarWgt | 0.76 | F | | | SHN_FxdBin | 4.65 | E | | | TRD_EquWgt | 7.74 | | G | | ENT_GenMax | 4.89 |

1. With no bias or small bias present, the inverse variance weighted (TRD_VarWgt) is still recommended as the most accurate estimator. The Generalized Maximum Entropy Estimator is the least accurate. SHN_Sheu's performance ranking is much lower than its competitors. Methods that use bins constructed with the overall or pooled standard deviation or even the fixed bin procedures (*_OStdBin, *_PStdBin, *_FxdBin) are recommended over SHN_Sheu method in this case. SHN_Sheu may be performing poorly because the lognormal distribution is skewed and Sheu's method does not distinguish the reliability of intervals on the left and on the right sides of the mean. When 15 inputs are used and 8 sources are sampled, the PAL_Sheu performs significantly better.

2. As the absolute value of the bias increases, SHN_Sheu's performance increases relatively under the condition that four sources are used. This pattern is reversed when eight sources are used, particularly with 15 inputs. When 15 inputs are used and 8 sources are sampled, SHN_Sheu's performance is the weakest as the bias is introduced. Clearly, the MAPE increases, in general, as the absolute value of the bias increases. The TRD_VarWgt's performance deteriorates as the absolute value of the bias increases. The use of Pal & Pal entropy does make the performance ranking of several methods better, although the performance is not necessarily significantly better.

3. The Generalized Maximum Entropy method (ENT_GenMax) is the best performer in the presence of large positive bias. The equal weight procedure (TRD_EquWgt) performs very well with the introduction of bias, in particular, in the case of 8 sources and 15 inputs.

4. Sensitivity analysis, conducted by increasing the number of inputs per data source, shows that the relative performance of the methods is altered substantially. The case of eight sources and 15 inputs shows that SHN_Sheu is not as robust as shown in the previous number of sources and inputs.

## 3.7. Conclusions for Data Fusion Methods Estimating the Population Mean

Essay 3 contributes to the academic literature by presenting new EBDF methods as well as presenting an estimation comparison under various distributions, biases, number of sources, and sample size (source size). This section addresses the research questions posed in Chapter 1 of this dissertation.

1. Can an alternative formulation for Shannon's entropy enhance the performance of Sheu's (2010)'s data fusion approach?

For a symmetric distribution, like the normal distribution, SHN_Sheu procedure is generally not significantly different in performance from the top performing data fusion methods in the presence of bias. However, if no bias is present, then the TRD_VarWgt is the clear choice. For a skewed distribution such as the lognormal, a procedure such as PAL_FxdBin or PAL_StdBin is recommended since its performance with 8 sources and 5 inputs from a lognormal distribution is among the top performing data fusion methods.

2. Do symmetric and skewed distributions affect the thirteen data fusion methods differently?

SHN_Sheu clearly performs better with inputs from a normal distribution and may perform quite poorly in the presence of many inputs from a lognormal distribution. Other procedures such as SHN_StdBin change rank ordering as seen in the configuration with 8 sources and 15 inputs from either a normal or lognormal distribution.

3. Do negative and positive biases affect the performance of the thirteen methods differently?

For many of the procedures, a positive or negative bias does not change the rank ordering performance of a data fusion method. However, the ENT_GenMax is easily affected by the bias and its performance depends on whether the bias is positive or negative.

4. Do entropy based data fusion outperform non-entropy based data fusion methods?

These entropy based data fusion methods perform better than the traditional methods when bias is present. In the presence of no bias, the traditional methods should be used.

5. Which data fusion methods are recommended for symmetric and skewed data sets when no bias is present?

Clearly the TRD_VarWgt method is a top performer in all scenarios without bias. In the case of no bias and 4 sources with 5 inputs, the top performing methods in addition to the TRD_VarWgt are SHN_FxdBin, SHN_PStdBin, and SHN_OStdBin. In the case of no bias and 4 sources with 15 inputs, the results are similar for both distributions. In the normal distribution case with no bias and 8 sources with 5 inputs, this same pattern holds and the TRD_VarWgt is significantly different from the others. In the lognormal distribution case with no bias and 8 sources with 5 inputs, SHN_FxdBin as well as PAL_FxdBin, PAL_OStdBin, and PAL_PStdBin are near the top performing method TRD_VarWgt. In the lognormal distribution case with no bias and 8 sources with 15 inputs, SHN_FxdBin and TRD_VarWgt are the top performing methods.

6. What is the recommendation under the condition of few data sources with bias?

The SHN_Sheu procedure is a strong contender when there are few data sources and bias is present. For positive bias, the ENT_GenMax is a top performing method.

# CHAPTER 4

# ESSAY 3: NEWSVENDOR MODEL: COMPARATIVE PERFORMANCE OF PROPOSED ENTROPY AND NON-ENTROPY BASED DATA FUSION ESTIMATORS

## 4.1. Abstract

This essay examines the effect of the data fusion methods proposed in Essay 2 on inventory optimization in the newsvendor model under the assumption that the demand is exponentially distributed. The exponential distribution is a distribution often used to model demand since demand is typically positive, unimodal, and somewhat skewed to the right. In addition, it is the generalized maximum entropy distribution when only the mean is known. For the newsvendor problem to have an optimal solution, the underlying distribution and its parameters need to be known, explicitly. In this essay, the underlying distribution is assumed to be known, but its parameter is not known. That is, demand has an exponential distribution with an unknown parameter $\lambda$. This essay tabulates the results for the 13 data fusion methods assessed in Essay 2. The basis for the comparison of the data fusion methods in Essay 3 is the profit using the optimal inventory level. The estimates from the data fusion methods provides the estimate for the mean, which in turn, is used to estimate the single parameter of the exponential distribution. The newsvendor problem illustrates an important SCM application, inventory optimization that the proposed EBDF methods may enhance. The configurations used in the Monte Carlo simulation in Essay 2 and the same configuration used in Essay 3 to construct MAPE performance measures for each data fusion method. This chapter illustrates the potential for EBDF methods in practical SCM applications.

## 4.2. Limited Knowledge of Demand Distribution in Newsvendor Applications

The newsvendor problem is instrumental to decision making in numerous SCM applications. A standard approach to the newsvendor problem is to simplify the estimation procedure by assuming full knowledge of the parametric distribution of the demand. In this case, solutions are straight-forward. However, this knowledge is not easily available. Andersson et al. (2013) considered the newsvendor problem under partial information. Because of limited data values and knowledge of the distribution, their study uses the entropy principles to assess the distribution of the demand data. However, an EBDF approach has not been fully investigated to assess the performance of estimators. Andersson et al. (2013) acknowledges the need for entropy-based methods: "To the best of our knowledge the operations management and revenue management literatures have not explored the use of maximum entropy methods to approximate unknown demand or willingness-to-pay distributions."

The classical newsvendor problem assumes that a merchant must sell a commodity in a market in which demand follows a probability distribution. The merchant orders a quantity q at a wholesale price and expects to make a profit. The merchant can sell the unsold items at a salvage price. The merchant would greatly benefit from knowing the optimal order quantity to maximize profit; however, knowledge about the distribution is often incomplete. If the distribution is known, but its parameters are unknown, then the merchant may decide to seek information about parameters such as the mean and the standard deviation, even if this information is subjective.

In this essay, the distribution of the demand will be assumed to be known, but its parameter must be estimated. If the distribution is exponential, then knowledge of the mean is sufficient to describe the distribution. The merchant may use (purchase) historical average sales from other merchants over several time periods. This process of collecting information from other vendors to deduce the demand can be viewed as a data fusion method in which the information from the vendor sources are "fused."

## 4.3. Description of Results of Newsvendor Model Performance for Methods

The results of Essay 2 in Chapter 3 are used to compute the optimal inventory in the newsvendor problem, which in turn provides the profit generated using that estimate. The relative performance of the data fusion methods is presented Tables 4.3.1 to 4.3.4 and 4.4.1 to 4.4.4. The same assumptions used in Essay 2 for the Monte Carlo Simulation study are repeated in this section. That is, the simulation study is replicated a 1000 times under the following conditions:

1. Normally or lognormally distributed inputs for the sources.
2. Eight Biases – four of which are negative and four of which are positive,
3. Two levels of number of sources – four or eight and
4. Levels of number of inputs to each source level – 5 or 15.

As in Essay 2, this essay is also assessing the sensitivity of the proposed data fusion methods to possible bias. As previously explained, this bias in added to the first source, which should be typically weighted the most since it is the most reliable. Adding the bias to less reliable sources will confound the effect of increased variance and bias. Both negative and positive bias effects are assessed since a method may tend to underestimate or overestimate a parameter of interest.

The same configurations as listed in Table 4.4.1 are again used in this study. That is, there are two levels of distributions – normal and lognormal, seven levels of biases – 3 negative, 3 positive, and one zero bias, two levels of sources – 4 sources and 8 sources, and two levels of sample size inputs – sample size of 5 and 15. Thus, a total of 56 configurations are used.

The first set of tables in this essay provides the relative performance for the profit estimation for the 13 data fusion procedures under the assumption that the inputs from the sources are normally distributed. The tables display the Tukey significance groupings by attaching the same letter next to procedures that are not significantly different.

## 4.4. Newsvendor Results from Simulation Study: Data Fusion Using Normally Distributed Inputs with Possible Bias

Using Andersson et al. (2013)'s notation, the newsvendor problem can be formulated as a maximization of profit: $\Pi(q,D) = (R-S)*\min(D,q) – (W-S)*q$, where q is the order quantity, D is the demand, W is the wholesale price, R is the revenue, and S is the salvage price. The expected value of the profit can be simplified as: $E(\Pi(q,D) ) = E[\min[D,q]] – \beta*q$, where $\beta$ is the ratio of (W-S)/(R-S) = (W-S)/(R-W + W-S). Overage cost is defined as W-S and underage cost is defined as R-W. The optimal order quantity is the solution to FD-1(1-$\beta$), where F is a known distribution of demand. For this simulation study, the value of $\beta$ is set to .95. This value implies that for an exponential distribution with parameter $\lambda$ = 1/1000, the optimal order quantity would be relatively small, 51. This optimal value was selected for this simulation since it is difficult to make a profit when the overage cost is close to the underage cost plus overage cost.

The simulation study in this essay uses the same configurations used in Essay 2. However, the resulting estimated mean from Essay 2 is used as input into solving the newsvendor problem for a known distribution, namely, the exponential distribution, with estimated parameter. Although the expectations are that the accuracy of the data fusion method methods in Essay 2 should result in a similar relative accuracy in Essay 3. This conclusion cannot be taken for granted since the variation of the estimates in Essay 2 will play a role in overage or underage costs that the merchant must pay. For the estimates from the data fusion methods in Essay 2 to result in order quantities close to the optimal order quantity, the estimates should be both accurate and precise, meaning, that the standard deviation of the estimate is small.

The error values (MAPE) listed in the Tables 4.3.1 to 4.3.4 and 4.4.1 to 4.4.4, are negative because they illustrate the deviation from the optimal inventory solution. The optimal solution to the newsvendor problem is known since the true population parameter is known in the simulation study. This allows for an accurate calculation of the MAPE. In practice, the true demand population parameters would be difficult to determine. The relative performances of the data fusion methods are tabulated in Tables 4.3.1 to 4.3.4 and 4.4.1 to 4.4.4.

As presented in Essay 2, eight tables were used to present the results. In this Essay, the tables will be organized in the same fashion. That is, Tables 4.3.1 through 4.3.4 assume that the inputs are normally distributed. The sources and inputs are 4 or 8 for the sources and 5 or 15 for the inputs. Tables 4.4.1 through 4.4.4 have similar configuration except that the distribution is the lognormal distribution. The methods listed near the top are the worst performing methods, whereas the methods listed at the bottom are the best performing methods. For each table, the first set of performance figures are under the assignment of no bias. The next three sets of performance measures are computed for the condition that bias is present and increases with the last set of performance measures illustrating the robustness of the methods under positive or negative bias with very large magnitude. The SHN_Sheu method is highlighted within the tables, just as was done in Essay 2. This method again serves as a bench mark for alternative data fusion methods to be compared.

Table 4.3.1 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 5 inputs from a normal distribution and assuming exponentially distributed demand.

| Normal Dist - Bias: None | | | | | | Normal Dist - Bias: Neg 100 | | | | | Normal Dist - Bias: Neg 300 | | | | | Normal Dist - Bias: Neg 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | | G | | ENT_GenMax | -0.13 | | E | | TRD_VarWgt | -0.50 | | G | | TRD_VarWgt | -4.05 | | G | TRD_VarWgt | -16.08 |
| | F | G | | SHN_StdBin | -0.13 | | | D | SHN_FxdBin | -0.27 | | F | | SHN_FxdBin | -1.89 | | F | SHN_FxdBin | -7.38 |
| E | F | G | | SHN_Sheu | -0.11 | | C | D | ENT_GenMax | -0.25 | | E | | SHN_PStdBin | -1.54 | | E | SHN_PStdBin | -5.96 |
| E | F | | D | PAL_StdBin | -0.11 | B | C | D | SHN_PStdBin | -0.24 | D | | | SHN_OStdBin | -1.05 | D | | PAL_FxdBin | -3.68 |
| E | F | C | D | PAL_Sheu | -0.11 | B | C | | SHN_OStdBin | -0.23 | D | C | | PAL_FxdBin | -0.98 | D | C | PAL_PStdBin | -3.36 |
| E | | C | D | TRD_EquWgt | -0.11 | B | A | | SHN_StdBin | -0.21 | B | C | | PAL_PStdBin | -0.91 | B | C | SHN_StdBin | -2.96 |
| B | | C | D | PAL_OStdBin | -0.09 | | A | | PAL_FxdBin | -0.19 | B | C | | ENT_GenMax | -0.86 | B | A | SHN_OStdBin | -2.72 |
| B | | C | D | PAL_PStdBin | -0.09 | | A | | PAL_PStdBin | -0.18 | B | | | SHN_StdBin | -0.83 | B | A | ENT_GenMax | -2.71 |
| B | | C | | PAL_FxdBin | -0.09 | | A | | PAL_OStdBin | -0.18 | B | A | | PAL_OStdBin | -0.79 | | A | PAL_OStdBin | -2.46 |
| B | | A | | SHN_OStdBin | -0.08 | | A | | SHN_Sheu | -0.18 | | A | | SHN_Sheu | -0.69 | | A | SHN_Sheu | -2.43 |
| B | | A | | SHN_PStdBin | -0.08 | | A | | PAL_StdBin | -0.17 | | A | | PAL_StdBin | -0.69 | | A | PAL_StdBin | -2.41 |
| B | | A | | SHN_FxdBin | -0.07 | | A | | PAL_Sheu | -0.17 | | A | | PAL_Sheu | -0.67 | | A | PAL_Sheu | -2.34 |
| | | A | | TRD_VarWgt | -0.06 | | A | | TRD_EquWgt | -0.17 | | A | | TRD_EquWgt | -0.67 | | A | TRD_EquWgt | -2.34 |

| Normal Dist - Bias: None | | | | | | Normal Dist - Bias: Pos 100 | | | | | Normal Dist - Bias: Pos 300 | | | | | Normal Dist - Bias: Pos 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | | G | | ENT_GenMax | -0.13 | | F | | TRD_VarWgt | -0.49 | | H | | TRD_VarWgt | -4.00 | | H | TRD_VarWgt | -15.80 |
| | F | G | | SHN_StdBin | -0.13 | | E | | SHN_FxdBin | -0.28 | | G | | SHN_FxdBin | -1.90 | | G | SHN_FxdBin | -7.35 |
| E | F | G | | SHN_Sheu | -0.11 | D | E | | SHN_PStdBin | -0.24 | | F | | SHN_PStdBin | -1.55 | | F | SHN_PStdBin | -5.93 |
| E | F | | D | PAL_StdBin | -0.11 | D | | | SHN_OStdBin | -0.23 | | | E | SHN_OStdBin | -1.03 | | E | PAL_FxdBin | -3.64 |
| E | F | C | D | PAL_Sheu | -0.11 | D | C | | SHN_StdBin | -0.21 | D | | E | PAL_FxdBin | -0.98 | D | E | PAL_PStdBin | -3.33 |
| E | | C | D | TRD_EquWgt | -0.11 | B | C | | PAL_FxdBin | -0.19 | D | C | E | PAL_PStdBin | -0.90 | D | C | SHN_StdBin | -2.98 |
| B | | C | D | PAL_OStdBin | -0.09 | B | C | | PAL_PStdBin | -0.18 | D | C | | SHN_StdBin | -0.85 | B | C | SHN_OStdBin | -2.69 |
| B | | C | D | PAL_PStdBin | -0.09 | B | C | | PAL_OStdBin | -0.18 | B | C | | PAL_OStdBin | -0.77 | B | A | PAL_OStdBin | -2.42 |
| B | | C | | PAL_FxdBin | -0.09 | B | C | | SHN_Sheu | -0.18 | B | | | SHN_Sheu | -0.69 | B | A | SHN_Sheu | -2.41 |
| B | | A | | SHN_OStdBin | -0.08 | B | A | | PAL_StdBin | -0.17 | B | | | PAL_StdBin | -0.68 | B | A | PAL_StdBin | -2.40 |
| B | | A | | SHN_PStdBin | -0.08 | B | A | | PAL_Sheu | -0.17 | B | | | PAL_Sheu | -0.66 | B | A | PAL_Sheu | -2.31 |
| B | | A | | SHN_FxdBin | -0.07 | B | A | | TRD_EquWgt | -0.17 | B | A | | TRD_EquWgt | -0.65 | B | A | TRD_EquWgt | -2.30 |
| | | A | | TRD_VarWgt | -0.06 | | A | | ENT_GenMax | -0.14 | | A | | ENT_GenMax | -0.52 | | A | ENT_GenMax | -2.01 |

Table 4.3.2 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 15 inputs from a normal distribution and assuming exponentially distributed demand.

| Normal Dist - Bias: None | | | Normal Dist - Bias: Neg 100 | | | Normal Dist - Bias: Neg 300 | | | Normal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| D | ENT_GenMax | -0.04 | E | TRD_VarWgt | -0.47 | I | TRD_VarWgt | -4.22 | H | TRD_VarWgt | -16.97 |
| D C | SHN_Sheu | -0.03 | D | SHN_FxdBin | -0.30 | H | SHN_FxdBin | -2.67 | G | SHN_FxdBin | -10.74 |
| C | SHN_StdBin | -0.03 | C | SHN_PStdBin | -0.27 | G | SHN_PStdBin | -2.34 | F | SHN_PStdBin | -9.40 |
| C | PAL_Sheu | -0.03 | C | SHN_OStdBin | -0.26 | F | SHN_OStdBin | -1.53 | E | PAL_FxdBin | -4.21 |
| C | TRD_EquWgt | -0.03 | B | PAL_FxdBin | -0.13 | E | PAL_FxdBin | -1.06 | D | PAL_PStdBin | -3.86 |
| C | PAL_StdBin | -0.03 | B | ENT_GenMax | -0.13 | D | PAL_PStdBin | -0.97 | C | SHN_OStdBin | -3.28 |
| B | PAL_OStdBin | -0.02 | B | PAL_PStdBin | -0.12 | C | PAL_OStdBin | -0.81 | B | PAL_OStdBin | -2.51 |
| B | PAL_PStdBin | -0.02 | B | PAL_OStdBin | -0.12 | B | ENT_GenMax | -0.69 | B A | ENT_GenMax | -2.46 |
| B | PAL_FxdBin | -0.02 | A | SHN_Sheu | -0.09 | A | SHN_Sheu | -0.59 | B A | SHN_Sheu | -2.31 |
| A | SHN_OStdBin | -0.01 | A | SHN_StdBin | -0.09 | A | SHN_StdBin | -0.58 | B A | SHN_StdBin | -2.24 |
| A | SHN_PStdBin | -0.01 | A | PAL_Sheu | -0.09 | A | PAL_Sheu | -0.57 | A | PAL_Sheu | -2.22 |
| A | SHN_FxdBin | -0.01 | A | TRD_EquWgt | -0.09 | A | TRD_EquWgt | -0.57 | A | TRD_EquWgt | -2.22 |
| A | TRD_VarWgt | -0.01 | A | PAL_StdBin | -0.09 | A | PAL_StdBin | -0.57 | A | PAL_StdBin | -2.21 |

| Normal Dist - Bias: None | | | Normal Dist - Bias: Pos 100 | | | Normal Dist - Bias: Pos 300 | | | Normal Dist - Bias: Pos 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| D | ENT_GenMax | -0.04 | F | TRD_VarWgt | -0.49 | I | TRD_VarWgt | -4.25 | I | TRD_VarWgt | -16.83 |
| D C | SHN_Sheu | -0.03 | E | SHN_FxdBin | -0.32 | H | SHN_FxdBin | -2.71 | H | SHN_FxdBin | -10.72 |
| C | SHN_StdBin | -0.03 | D | SHN_PStdBin | -0.28 | G | SHN_PStdBin | -2.38 | G | SHN_PStdBin | -9.40 |
| C | PAL_Sheu | -0.03 | D | SHN_OStdBin | -0.27 | F | SHN_OStdBin | -1.56 | F | PAL_FxdBin | -4.26 |
| C | TRD_EquWgt | -0.03 | C | PAL_FxdBin | -0.15 | E | PAL_FxdBin | -1.09 | E | PAL_PStdBin | -3.91 |
| C | PAL_StdBin | -0.03 | C | PAL_PStdBin | -0.14 | D | PAL_PStdBin | -1.01 | D | SHN_OStdBin | -3.33 |
| B | PAL_OStdBin | -0.02 | C | PAL_OStdBin | -0.13 | C | PAL_OStdBin | -0.84 | C | PAL_OStdBin | -2.56 |
| B | PAL_PStdBin | -0.02 | B | SHN_Sheu | -0.10 | B | SHN_Sheu | -0.62 | B C | SHN_Sheu | -2.36 |
| B | PAL_FxdBin | -0.02 | B | SHN_StdBin | -0.10 | B | SHN_StdBin | -0.61 | B A C | SHN_StdBin | -2.30 |
| A | SHN_OStdBin | -0.01 | B | PAL_Sheu | -0.10 | B | PAL_Sheu | -0.60 | B A | PAL_Sheu | -2.27 |
| A | SHN_PStdBin | -0.01 | B | TRD_EquWgt | -0.10 | B | TRD_EquWgt | -0.60 | B A | TRD_EquWgt | -2.27 |
| A | SHN_FxdBin | -0.01 | B | PAL_StdBin | -0.10 | B | PAL_StdBin | -0.60 | B A | PAL_StdBin | -2.27 |
| A | TRD_VarWgt | -0.01 | A | ENT_GenMax | -0.07 | A | ENT_GenMax | -0.49 | A | ENT_GenMax | -2.05 |

Table 4.3.3 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 5 inputs from a normal distribution and assuming exponentially distributed demand.

**Normal Dist - Bias: None**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | G | | ENT_GenMax | -0.27 |
| | G | | SHN_StdBin | -0.25 |
| F | | | SHN_Sheu | -0.19 |
| F | | | PAL_StdBin | -0.18 |
| F | | E | PAL_Sheu | -0.18 |
| F | D | E | TRD_EquWgt | -0.17 |
| C | D | E | PAL_PStdBin | -0.14 |
| C | D | E | PAL_OStdBin | -0.14 |
| C | D | | PAL_FxdBin | -0.14 |
| C | B | | SHN_OStdBin | -0.13 |
| C | B | | SHN_PStdBin | -0.13 |
| | B | | SHN_FxdBin | -0.10 |
| | A | | TRD_VarWgt | -0.06 |

**Normal Dist - Bias: Neg 100**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| D | | | TRD_VarWgt | -0.43 |
| C | | | ENT_GenMax | -0.35 |
| B | | | SHN_StdBin | -0.26 |
| A | | | SHN_Sheu | -0.20 |
| A | | | PAL_StdBin | -0.19 |
| A | | | PAL_Sheu | -0.19 |
| A | | | TRD_EquWgt | -0.18 |
| A | | | SHN_FxdBin | -0.18 |
| A | | | PAL_FxdBin | -0.16 |
| A | | | SHN_PStdBin | -0.16 |
| A | | | PAL_PStdBin | -0.16 |
| A | | | PAL_OStdBin | -0.16 |
| A | | | SHN_OStdBin | -0.16 |

**Normal Dist - Bias: Neg 300**

| Pair-wise Signif | | | | Method | MAPE (%) |
|---|---|---|---|---|---|
| | | G | | TRD_VarWgt | -3.42 |
| | | F | | SHN_FxdBin | -0.91 |
| | | E | | ENT_GenMax | -0.61 |
| | D | | | SHN_PStdBin | -0.45 |
| | D | | C | SHN_StdBin | -0.43 |
| B | D | | C | SHN_OStdBin | -0.42 |
| B | D | A | C | PAL_FxdBin | -0.40 |
| B | | A | C | PAL_PStdBin | -0.33 |
| B | | A | C | PAL_OStdBin | -0.32 |
| B | | A | | SHN_Sheu | -0.32 |
| | | A | | PAL_StdBin | -0.31 |
| | | A | | PAL_Sheu | -0.30 |
| | | A | | TRD_EquWgt | -0.30 |

**Normal Dist - Bias: Neg 600**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | F | | TRD_VarWgt | -13.54 |
| | E | | SHN_FxdBin | -3.44 |
| | D | | SHN_PStdBin | -1.45 |
| | D | C | ENT_GenMax | -1.23 |
| | D | C | PAL_FxdBin | -1.23 |
| B | D | C | SHN_OStdBin | -1.12 |
| B | A | C | SHN_StdBin | -1.01 |
| B | A | C | PAL_PStdBin | -0.91 |
| B | A | | PAL_OStdBin | -0.82 |
| | A | | SHN_Sheu | -0.74 |
| | A | | PAL_StdBin | -0.72 |
| | A | | PAL_Sheu | -0.70 |
| | A | | TRD_EquWgt | -0.69 |

**Normal Dist - Bias: None**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | G | | ENT_GenMax | -0.27 |
| | G | | SHN_StdBin | -0.25 |
| F | | | SHN_Sheu | -0.19 |
| F | | | PAL_StdBin | -0.18 |
| F | | E | PAL_Sheu | -0.18 |
| F | D | E | TRD_EquWgt | -0.17 |
| C | D | E | PAL_PStdBin | -0.14 |
| C | D | E | PAL_OStdBin | -0.14 |
| C | D | | PAL_FxdBin | -0.14 |
| C | B | | SHN_OStdBin | -0.13 |
| C | B | | SHN_PStdBin | -0.13 |
| | B | | SHN_FxdBin | -0.10 |
| | A | | TRD_VarWgt | -0.06 |

**Normal Dist - Bias: Pos 100**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | E | | TRD_VarWgt | -0.43 |
| | D | | SHN_StdBin | -0.28 |
| | | C | ENT_GenMax | -0.21 |
| B | | C | SHN_Sheu | -0.21 |
| B | A | C | PAL_StdBin | -0.20 |
| B | A | C | SHN_FxdBin | -0.20 |
| B | A | C | PAL_Sheu | -0.20 |
| B | A | C | TRD_EquWgt | -0.19 |
| B | A | C | PAL_FxdBin | -0.18 |
| B | A | | PAL_OStdBin | -0.17 |
| B | A | | PAL_PStdBin | -0.17 |
| B | A | | SHN_OStdBin | -0.17 |
| | A | | SHN_PStdBin | -0.17 |

**Normal Dist - Bias: Pos 300**

| Pair-wise Signif | | Method | MAPE (%) |
|---|---|---|---|
| | F | TRD_VarWgt | -3.39 |
| | E | SHN_FxdBin | -0.98 |
| | D | SHN_StdBin | -0.48 |
| | D | SHN_PStdBin | -0.47 |
| C | D | SHN_OStdBin | -0.45 |
| C | D | PAL_FxdBin | -0.45 |
| C | B | PAL_PStdBin | -0.36 |
| C | B | SHN_Sheu | -0.35 |
| C | B | PAL_OStdBin | -0.35 |
| C | B | PAL_StdBin | -0.35 |
| | B | PAL_Sheu | -0.33 |
| | B | TRD_EquWgt | -0.33 |
| | A | ENT_GenMax | -0.20 |

**Normal Dist - Bias: Pos 600**

| Pair-wise Signif | | | Method | MAPE (%) |
|---|---|---|---|---|
| | G | | TRD_VarWgt | -13.34 |
| | F | | SHN_FxdBin | -3.56 |
| E | | | SHN_PStdBin | -1.49 |
| E | | D | PAL_FxdBin | -1.32 |
| E | C | D | SHN_OStdBin | -1.16 |
| B | C | D | SHN_StdBin | -1.11 |
| B | C | D | PAL_PStdBin | -0.97 |
| B | C | | PAL_OStdBin | -0.88 |
| B | C | | SHN_Sheu | -0.81 |
| B | | | PAL_StdBin | -0.80 |
| B | A | | PAL_Sheu | -0.77 |
| B | A | | TRD_EquWgt | -0.76 |
| | A | | ENT_GenMax | -0.41 |

Table 4.3.4 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 15 inputs from a normal distribution and assuming exponentially distributed demand.

| Normal Dist - Bias: None | | | Normal Dist - Bias: Neg 100 | | | Normal Dist - Bias: Neg 300 | | | Normal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| E | ENT_GenMax | -0.10 | D | TRD_VarWgt | -0.41 | F | TRD_VarWgt | -3.66 | F | TRD_VarWgt | -14.69 |
| D | SHN_Sheu | -0.05 | C | SHN_FxdBin | -0.18 | E | SHN_FxdBin | -1.55 | E | SHN_FxdBin | -6.22 |
| D | SHN_StdBin | -0.05 | C | ENT_GenMax | -0.17 | D | SHN_PStdBin | -0.59 | D | SHN_PStdBin | -2.34 |
| D | PAL_Sheu | -0.05 | B | SHN_PStdBin | -0.08 | D | SHN_OStdBin | -0.53 | C | SHN_OStdBin | -1.62 |
| D | TRD_EquWgt | -0.05 | B | SHN_OStdBin | -0.08 | C | ENT_GenMax | -0.40 | C | PAL_FxdBin | -1.44 |
| D | PAL_StdBin | -0.05 | B A | PAL_FxdBin | -0.07 | C | PAL_FxdBin | -0.38 | B | PAL_PStdBin | -0.97 |
| C | PAL_FxdBin | -0.04 | B A | SHN_Sheu | -0.07 | B | PAL_PStdBin | -0.26 | B | ENT_GenMax | -0.97 |
| C | PAL_OStdBin | -0.04 | A | SHN_StdBin | -0.06 | B | PAL_OStdBin | -0.25 | B | PAL_OStdBin | -0.84 |
| C | PAL_PStdBin | -0.04 | A | PAL_Sheu | -0.06 | A | SHN_Sheu | -0.19 | A | SHN_Sheu | -0.62 |
| B | SHN_OStdBin | -0.02 | A | TRD_EquWgt | -0.06 | A | SHN_StdBin | -0.18 | A | SHN_StdBin | -0.59 |
| B A | SHN_PStdBin | -0.02 | A | PAL_StdBin | -0.06 | A | TRD_EquWgt | -0.18 | A | TRD_EquWgt | -0.58 |
| B A | SHN_FxdBin | -0.02 | A | PAL_OStdBin | -0.06 | A | PAL_Sheu | -0.18 | A | PAL_Sheu | -0.58 |
| A | TRD_VarWgt | -0.01 | A | PAL_PStdBin | -0.06 | A | PAL_StdBin | -0.18 | A | PAL_StdBin | -0.58 |

| Normal Dist - Bias: None | | | Normal Dist - Bias: Pos 100 | | | Normal Dist - Bias: Pos 300 | | | Normal Dist - Bias: Pos 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| E | ENT_GenMax | -0.10 | E | TRD_VarWgt | -0.43 | H | TRD_VarWgt | -3.68 | G | TRD_VarWgt | -14.60 |
| D | SHN_Sheu | -0.05 | D | SHN_FxdBin | -0.20 | G | SHN_FxdBin | -1.60 | F | SHN_FxdBin | -6.27 |
| D | SHN_StdBin | -0.05 | C | SHN_PStdBin | -0.09 | F | SHN_PStdBin | -0.62 | E | SHN_PStdBin | -2.39 |
| D | PAL_Sheu | -0.05 | C | SHN_OStdBin | -0.09 | E | SHN_OStdBin | -0.56 | D | SHN_OStdBin | -1.67 |
| D | TRD_EquWgt | -0.05 | B C | PAL_FxdBin | -0.08 | D | PAL_FxdBin | -0.41 | D | PAL_FxdBin | -1.50 |
| D | PAL_StdBin | -0.05 | B A | SHN_Sheu | -0.07 | C | PAL_PStdBin | -0.29 | C | PAL_PStdBin | -1.02 |
| C | PAL_FxdBin | -0.04 | B A | SHN_StdBin | -0.07 | C | PAL_OStdBin | -0.28 | C | PAL_OStdBin | -0.89 |
| C | PAL_OStdBin | -0.04 | B A | PAL_Sheu | -0.07 | B | SHN_Sheu | -0.21 | B | SHN_Sheu | -0.67 |
| C | PAL_PStdBin | -0.04 | B A | PAL_StdBin | -0.07 | B | SHN_StdBin | -0.20 | B | SHN_StdBin | -0.64 |
| B | SHN_OStdBin | -0.02 | B A | TRD_EquWgt | -0.07 | B | PAL_Sheu | -0.20 | B | PAL_Sheu | -0.63 |
| B A | SHN_PStdBin | -0.02 | A | PAL_PStdBin | -0.07 | B | PAL_StdBin | -0.20 | B | PAL_StdBin | -0.63 |
| B A | SHN_FxdBin | -0.02 | A | PAL_OStdBin | -0.07 | B | TRD_EquWgt | -0.20 | B | TRD_EquWgt | -0.63 |
| A | TRD_VarWgt | -0.01 | A | ENT_GenMax | -0.07 | A | ENT_GenMax | -0.08 | A | ENT_GenMax | -0.34 |

What do the simulation results under the assumption of normally distributed responses reveal about the performance of the 13 data fusion methods in estimating the optimal inventory level? The expectation is that the performance of these methods in the newsvendor problem may parallel the results from the previous essay in estimating the optimal inventory value. However, that pattern does not need to be the same, since there are now costs involved. On an absolute basis, the MAPE is small and is less than 1%. This is due to the choice of cost ratios selected for the newsvendor problem. The tables reveal several patterns for the performance of the data fusion methods. Although it is true that changing the configuration values will change the ordering of the methods' performance, the patterns are worth noting as they provide information about the consistency of a method's performance. The following results are deduced from these four tables:

1. With the normally distributed data, the performance of data fusion methods used in the newsvendor problem mimics the order of the performance in the estimation of the population mean parameter. Some significance groupings change, but most of the methods maintain their same rank ordering.

2. SHN_Sheu's performance is relatively stronger in estimating the population mean parameter than in solving the newsvendor problem. In the configuration with 4 sources and either 5 or 15 inputs from the normal distribution, SHN_Sheu's performance is relatively weaker under the newsvendor problem results especially for the case in which the magnitude of the bias is higher.

3. Pal_StdBin's relative performance improves slightly under the newsvendor problem with the introduction of bias under the scenario with 8 sources and 15 inputs from the normal distribution.

4. The traditional equal weight (TRD_EquWgt) data fusion method is the best or second to best for the newsvendor problem with four sources and 5 inputs from the normal distribution with the introduction of bias, a somewhat stronger performance than in estimating the population mean.

## 4.5. Newsvendor Results from Simulation Study: Data Fusion Using Lognormally Distributed Inputs with Possible Bias

A review of the data fusion methods in Essay 2 for the lognormal distribution shows that SHN_Sheu was a strong performer with the introduction of bias for the scenarios with 4 sources and 5 or 15 inputs from the normal population. With a larger number of sources, namely 8, from with lognormal inputs, SHN_Sheu did not perform as strongly in estimating the mean. The results for the data fusion methods will now be examined. The following results are deduced from Tables 4.4.1-4.4.4.

The lognormal distribution is a skewed distribution and therefore, this distribution shape, potentially allows alternative methods to Sheu (2010)'s originally formulation to be viable challengers. For example, PAL_StdBin appears to be performing at least as well as SHN_Sheu. The fact that formulations such as the ones that use the StdBin allow for information to be captured from both sides of the mean separately allows for asymmetric information to be utilized.

As mentioned previously, a strong performing data fusion method to approximate the mean may not necessarily have as strong performance for the newsvendor model. The reason is that an estimator that is close to the true estimate, but perhaps errs mostly by overestimating the parameter may contribute to higher costs. The value in assessing the parameter estimates of the methods is that a cost is now used to determine how well the methods perform.

The data fusion methods that are either consistently at the top or bottom are noteworthy. These results shed light on the importance of the simulation configuration. For example, SHN_Sheu performs well with few sources, but not with a large number when assuming a lognormal distribution.

Table 4.4.1 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 5 inputs from a lognormal distribution and assuming exponentially distributed demand.

| Lognormal Dist - Bias: None | | | | Lognormal Dist - Bias: Neg 100 | | | | Lognormal Dist - Bias: Neg 300 | | | | Lognormal Dist - Bias: Neg 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | E | SHN_StdBin | -0.09 | | D | TRD_VarWgt | -0.54 | | F | TRD_VarWgt | -5.48 | | E | TRD_VarWgt | -29.09 |
| D | E | ENT_GenMax | -0.08 | | C | SHN_FxdBin | -0.22 | | E | SHN_FxdBin | -1.56 | | D | SHN_FxdBin | -6.09 |
| D | E | SHN_Sheu | -0.07 | | C | SHN_PStdBin | -0.22 | | E | SHN_PStdBin | -1.55 | | D | SHN_PStdBin | -6.05 |
| D | | PAL_StdBin | -0.07 | | C | SHN_OStdBin | -0.20 | D | | SHN_OStdBin | -0.92 | | C | PAL_FxdBin | -3.38 |
| D | C | PAL_Sheu | -0.07 | | C | ENT_GenMax | -0.19 | D | C | PAL_FxdBin | -0.89 | | C | PAL_PStdBin | -3.37 |
| D | C | TRD_EquWgt | -0.07 | B | | SHN_StdBin | -0.16 | D | C | PAL_PStdBin | -0.89 | B | | SHN_StdBin | -2.91 |
| B | C | PAL_OStdBin | -0.06 | B | A | PAL_PStdBin | -0.15 | B | C | SHN_StdBin | -0.79 | B | A | ENT_GenMax | -2.59 |
| B | C | PAL_FxdBin | -0.06 | B | A | PAL_FxdBin | -0.15 | B | C | ENT_GenMax | -0.78 | | A | SHN_OStdBin | -2.49 |
| B | C | PAL_PStdBin | -0.06 | B | A | PAL_OStdBin | -0.15 | B | A | PAL_OStdBin | -0.72 | | A | SHN_Sheu | -2.39 |
| B | A | SHN_OStdBin | -0.05 | B | A | SHN_Sheu | -0.14 | | A | SHN_Sheu | -0.65 | | A | PAL_StdBin | -2.37 |
| B | A | SHN_FxdBin | -0.05 | B | A | PAL_StdBin | -0.14 | | A | PAL_StdBin | -0.64 | | A | PAL_OStdBin | -2.35 |
| B | A | SHN_PStdBin | -0.05 | | A | PAL_Sheu | -0.13 | | A | PAL_Sheu | -0.63 | | A | PAL_Sheu | -2.30 |
| | A | TRD_VarWgt | -0.04 | | A | TRD_EquWgt | -0.13 | | A | TRD_EquWgt | -0.63 | | A | TRD_EquWgt | -2.29 |

| Lognormal Dist - Bias: None | | | | Lognormal Dist - Bias: Pos 111 | | | | Lognormal Dist - Bias: Pos 428 | | | | Lognormal Dist - Bias: Pos 1500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | E | SHN_StdBin | -0.09 | | F | TRD_VarWgt | -0.50 | | F | TRD_VarWgt | -5.12 | | E | TRD_VarWgt | -23.91 |
| D | E | ENT_GenMax | -0.08 | E | | SHN_FxdBin | -0.24 | | E | SHN_FxdBin | -2.79 | | D | SHN_PStdBin | -20.24 |
| D | E | SHN_Sheu | -0.07 | E | | SHN_PStdBin | -0.23 | | E | SHN_PStdBin | -2.63 | | D | SHN_FxdBin | -19.85 |
| D | | PAL_StdBin | -0.07 | E | D | SHN_OStdBin | -0.21 | D | | PAL_FxdBin | -1.65 | | C | SHN_StdBin | -17.66 |
| D | C | PAL_Sheu | -0.07 | C | D | SHN_StdBin | -0.19 | D | | PAL_PStdBin | -1.61 | B | | PAL_PStdBin | -15.50 |
| D | C | TRD_EquWgt | -0.07 | C | B | PAL_FxdBin | -0.17 | D | | SHN_StdBin | -1.54 | B | | PAL_FxdBin | -15.26 |
| B | C | PAL_OStdBin | -0.06 | C | B | PAL_PStdBin | -0.16 | D | C | SHN_OStdBin | -1.43 | B | A | SHN_Sheu | -14.47 |
| B | C | PAL_FxdBin | -0.06 | C | B | PAL_OStdBin | -0.16 | B | C | PAL_OStdBin | -1.27 | B | A | PAL_StdBin | -14.32 |
| B | C | PAL_PStdBin | -0.06 | | B | SHN_Sheu | -0.15 | B | C | SHN_Sheu | -1.26 | B | A | SHN_OStdBin | -14.16 |
| B | A | SHN_OStdBin | -0.05 | | B | PAL_StdBin | -0.15 | B | C | PAL_StdBin | -1.24 | B | A | PAL_Sheu | -13.88 |
| B | A | SHN_FxdBin | -0.05 | | B | PAL_Sheu | -0.15 | B | A | PAL_Sheu | -1.20 | B | A | PAL_OStdBin | -13.86 |
| B | A | SHN_PStdBin | -0.05 | | B | TRD_EquWgt | -0.14 | B | A | TRD_EquWgt | -1.19 | B | A | TRD_EquWgt | -13.80 |
| | A | TRD_VarWgt | -0.04 | | A | ENT_GenMax | -0.10 | | A | ENT_GenMax | -1.00 | | A | ENT_GenMax | -13.07 |

Table 4.4.2 Mean Absolute Percent Error in estimation optimal inventory using 4 sources with 15 inputs from a lognormal distribution and assuming exponentially distributed demand.

| Lognormal Dist - Bias: None | | | Lognormal Dist - Bias: Neg 100 | | | Lognormal Dist - Bias: Neg 300 | | | Lognormal Dist - Bias: Neg 600 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| D | ENT_GenMax | -0.03 | F | TRD_VarWgt | -0.53 | G | TRD_VarWgt | -5.86 | G | TRD_VarWgt | -5.86 |
| D | SHN_Sheu | -0.02 | E | SHN_FxdBin | -0.28 | F | SHN_FxdBin | -2.56 | F | SHN_FxdBin | -2.56 |
| C | SHN_StdBin | -0.02 | D | SHN_PStdBin | -0.26 | E | SHN_PStdBin | -2.36 | E | SHN_PStdBin | -2.36 |
| C | PAL_Sheu | -0.02 | C | SHN_OStdBin | -0.24 | D | SHN_OStdBin | -1.25 | D | SHN_OStdBin | -1.25 |
| C | PAL_StdBin | -0.02 | B | ENT_GenMax | -0.12 | C | PAL_FxdBin | -1.01 | C | PAL_FxdBin | -1.01 |
| C | TRD_EquWgt | -0.02 | B | PAL_FxdBin | -0.12 | C | PAL_PStdBin | -0.97 | C | PAL_PStdBin | -0.97 |
| B | PAL_OStdBin | -0.01 | B | PAL_PStdBin | -0.12 | B | PAL_OStdBin | -0.75 | B | PAL_OStdBin | -0.75 |
| B | PAL_PStdBin | -0.01 | B | PAL_OStdBin | -0.11 | B | ENT_GenMax | -0.68 | B | ENT_GenMax | -0.68 |
| B | PAL_FxdBin | -0.01 | A | SHN_Sheu | -0.09 | A | SHN_Sheu | -0.61 | A | SHN_Sheu | -0.61 |
| A | SHN_OStdBin | -0.01 | A | SHN_StdBin | -0.08 | A | SHN_StdBin | -0.56 | A | SHN_StdBin | -0.56 |
| A | SHN_PStdBin | -0.01 | A | PAL_Sheu | -0.08 | A | TRD_EquWgt | -0.56 | A | TRD_EquWgt | -0.56 |
| A | SHN_FxdBin | -0.01 | A | TRD_EquWgt | -0.08 | A | PAL_Sheu | -0.56 | A | PAL_Sheu | -0.56 |
| A | TRD_VarWgt | -0.01 | A | PAL_StdBin | -0.08 | A | PAL_StdBin | -0.56 | A | PAL_StdBin | -0.56 |

| Lognormal Dist - Bias: None | | | Lognormal Dist - Bias: Pos 111 | | | Lognormal Dist - Bias: Pos 428 | | | Lognormal Dist - Bias: Pos 1500 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) |
| D | ENT_GenMax | -0.03 | G | TRD_VarWgt | -0.51 | F | TRD_VarWgt | -5.09 | F | TRD_VarWgt | -18.20 |
| D | SHN_Sheu | -0.02 | F | SHN_FxdBin | -0.33 | E | SHN_FxdBin | -3.65 | E | SHN_PStdBin | -17.09 |
| C | SHN_StdBin | -0.02 | E | SHN_PStdBin | -0.31 | E | SHN_PStdBin | -3.60 | E | SHN_FxdBin | -16.26 |
| C | PAL_Sheu | -0.02 | D | SHN_OStdBin | -0.28 | D | SHN_OStdBin | -1.86 | D | PAL_PStdBin | -15.12 |
| C | PAL_StdBin | -0.02 | C | PAL_FxdBin | -0.15 | D | PAL_FxdBin | -1.82 | D C | SHN_Sheu | -14.90 |
| C | TRD_EquWgt | -0.02 | C | PAL_PStdBin | -0.15 | D | PAL_PStdBin | -1.80 | B D C | PAL_FxdBin | -14.77 |
| B | PAL_OStdBin | -0.01 | C | PAL_OStdBin | -0.14 | C | PAL_OStdBin | -1.36 | B A C | TRD_EquWgt | -13.83 |
| B | PAL_PStdBin | -0.01 | B | SHN_Sheu | -0.11 | C B | SHN_Sheu | -1.26 | B A C | SHN_OStdBin | -13.81 |
| B | PAL_FxdBin | -0.01 | B | SHN_StdBin | -0.10 | B | TRD_EquWgt | -1.17 | B A C | PAL_OStdBin | -13.81 |
| A | SHN_OStdBin | -0.01 | B | TRD_EquWgt | -0.10 | B | SHN_StdBin | -1.16 | B A C | SHN_StdBin | -13.80 |
| A | SHN_PStdBin | -0.01 | B | PAL_Sheu | -0.10 | B | PAL_Sheu | -1.16 | B A | PAL_Sheu | -13.74 |
| A | SHN_FxdBin | -0.01 | B | PAL_StdBin | -0.10 | B | PAL_StdBin | -1.16 | B A | PAL_StdBin | -13.72 |
| A | TRD_VarWgt | -0.01 | A | ENT_GenMax | -0.06 | A | ENT_GenMax | -1.00 | A | ENT_GenMax | -13.23 |

Table 4.4.3 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 5 inputs from a lognormal distribution and assuming exponentially distributed demand.

| Lognormal Dist - Bias: None | | | | Lognormal Dist - Bias: Neg 100 | | | | Lognormal Dist - Bias: Neg 300 | | | | Lognormal Dist - Bias: Neg 600 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | C | SHN_StdBin | -32.04 | | C | SHN_StdBin | -31.95 | B | | SHN_StdBin | -31.87 | B | | SHN_StdBin | -32.00 |
| B | | SHN_FxdBin | -15.84 | B | | SHN_FxdBin | -15.87 | A | | SHN_FxdBin | -16.47 | B | | TRD_VarWgt | -28.88 |
| B | A | PAL_StdBin | -10.80 | B | A | PAL_StdBin | -10.76 | A | | PAL_StdBin | -10.75 | B | A | SHN_FxdBin | -18.64 |
| B | A | SHN_Sheu | -8.09 | B | A | SHN_Sheu | -8.07 | A | | SHN_Sheu | -8.10 | | A | PAL_StdBin | -10.96 |
| B | A | PAL_FxdBin | -7.38 | B | A | PAL_FxdBin | -7.38 | A | | PAL_FxdBin | -7.57 | | A | SHN_Sheu | -8.37 |
| B | A | SHN_OStdBin | -6.88 | B | A | SHN_OStdBin | -6.95 | A | | SHN_OStdBin | -7.25 | | A | PAL_FxdBin | -8.32 |
| B | A | SHN_PStdBin | -6.85 | B | A | SHN_PStdBin | -6.93 | A | | SHN_PStdBin | -7.25 | | A | SHN_PStdBin | -8.07 |
| B | A | PAL_Sheu | -6.26 | B | A | PAL_Sheu | -6.24 | A | | PAL_Sheu | -6.30 | | A | SHN_OStdBin | -8.01 |
| B | A | PAL_OStdBin | -5.49 | B | A | PAL_OStdBin | -5.52 | A | | PAL_PStdBin | -5.69 | | A | PAL_Sheu | -6.60 |
| B | A | PAL_PStdBin | -5.48 | B | A | PAL_PStdBin | -5.52 | A | | PAL_OStdBin | -5.68 | | A | PAL_PStdBin | -6.23 |
| B | A | TRD_EquWgt | -5.00 | B | A | TRD_EquWgt | -4.99 | A | | TRD_VarWgt | -5.63 | | A | PAL_OStdBin | -6.20 |
| B | A | ENT_GenMax | -2.43 | B | A | ENT_GenMax | -2.73 | A | | TRD_EquWgt | -5.07 | | A | TRD_EquWgt | -5.42 |
| | A | TRD_VarWgt | -0.15 | | A | TRD_VarWgt | -0.74 | A | | ENT_GenMax | -3.42 | | A | ENT_GenMax | -4.69 |

| Lognormal Dist - Bias: None | | | | Lognormal Dist - Bias: Pos 111 | | | | Lognormal Dist - Bias: Pos 428 | | | | Lognormal Dist - Bias: Pos 1500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) | Pair-wise Signif | | Method | MAPE (%) |
| | C | SHN_StdBin | -32.04 | | C | SHN_StdBin | -32.17 | | C | SHN_StdBin | -32.75 | | C | SHN_StdBin | -37.04 |
| B | | SHN_FxdBin | -15.84 | B | | SHN_FxdBin | -16.00 | B | C | SHN_FxdBin | -18.36 | B | C | SHN_FxdBin | -30.85 |
| B | A | PAL_StdBin | -10.80 | B | A | PAL_StdBin | -10.88 | B | A | PAL_StdBin | -11.31 | B | A | TRD_VarWgt | -19.08 |
| B | A | SHN_Sheu | -8.09 | B | A | SHN_Sheu | -8.15 | B | A | SHN_Sheu | -8.53 | B | A | PAL_StdBin | -14.83 |
| B | A | PAL_FxdBin | -7.38 | B | A | PAL_FxdBin | -7.45 | B | A | PAL_FxdBin | -8.14 | B | A | PAL_FxdBin | -13.41 |
| B | A | SHN_OStdBin | -6.88 | B | A | SHN_OStdBin | -6.85 | B | A | SHN_OStdBin | -7.06 | | A | SHN_Sheu | -11.91 |
| B | A | SHN_PStdBin | -6.85 | B | A | SHN_PStdBin | -6.80 | B | A | SHN_PStdBin | -7.01 | | A | SHN_PStdBin | -11.20 |
| B | A | PAL_Sheu | -6.26 | B | A | PAL_Sheu | -6.31 | B | A | PAL_Sheu | -6.66 | | A | SHN_OStdBin | -10.29 |
| B | A | PAL_OStdBin | -5.49 | B | A | PAL_OStdBin | -5.50 | B | A | PAL_OStdBin | -5.78 | | A | PAL_Sheu | -10.01 |
| B | A | PAL_PStdBin | -5.48 | B | A | PAL_PStdBin | -5.49 | B | A | PAL_PStdBin | -5.76 | | A | PAL_PStdBin | -9.36 |
| B | A | TRD_EquWgt | -5.00 | B | A | TRD_EquWgt | -5.04 | B | A | TRD_EquWgt | -5.38 | | A | PAL_OStdBin | -9.10 |
| B | A | ENT_GenMax | -2.43 | B | A | ENT_GenMax | -2.13 | B | A | TRD_VarWgt | -4.21 | | A | TRD_EquWgt | -8.78 |
| | A | TRD_VarWgt | -0.15 | | A | TRD_VarWgt | -0.44 | | A | ENT_GenMax | -1.50 | | A | ENT_GenMax | -1.63 |

Table 4.4.3 Mean Absolute Percent Error in estimation optimal inventory using 8 sources with 15 inputs from a lognormal distribution and assuming exponentially distributed demand.

| Lognormal Dist - Bias: None | | | Lognormal Dist - Bias: Neg 100 | | | Lognormal Dist - Bias: Neg 300 | | | Lognormal Dist - Bias: Neg 600 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) |
| C | SHN_Sheu | -33.75 | C | SHN_Sheu | -33.76 | C | SHN_Sheu | -33.80 | | C | | SHN_Sheu | -33.95 |
| B | SHN_StdBin | -14.70 | B | SHN_StdBin | -14.67 | B | SHN_StdBin | -14.66 | | C | | TRD_VarWgt | -30.49 |
| A | SHN_PStdBin | -4.17 | A | SHN_PStdBin | -4.21 | A | TRD_VarWgt | -5.79 | B | | | SHN_StdBin | -14.76 |
| A | SHN_OStdBin | -4.16 | A | SHN_OStdBin | -4.20 | A | SHN_PStdBin | -4.45 | B | A | | SHN_FxdBin | -7.16 |
| A | PAL_StdBin | -3.51 | A | PAL_StdBin | -3.49 | A | SHN_OStdBin | -4.45 | | A | | SHN_PStdBin | -5.22 |
| A | PAL_Sheu | -2.49 | A | ENT_GenMax | -2.53 | A | PAL_StdBin | -3.52 | | A | | SHN_OStdBin | -5.20 |
| A | ENT_GenMax | -2.17 | A | PAL_Sheu | -2.48 | A | ENT_GenMax | -3.36 | | A | | ENT_GenMax | -4.83 |
| A | PAL_OStdBin | -1.73 | A | PAL_OStdBin | -1.76 | A | PAL_Sheu | -2.54 | | A | | PAL_StdBin | -3.75 |
| A | PAL_PStdBin | -1.73 | A | PAL_PStdBin | -1.76 | A | SHN_FxdBin | -2.00 | | A | | PAL_Sheu | -2.82 |
| A | TRD_EquWgt | -1.19 | A | TRD_EquWgt | -1.20 | A | PAL_OStdBin | -1.93 | | A | | PAL_OStdBin | -2.44 |
| A | PAL_FxdBin | -0.79 | A | PAL_FxdBin | -0.83 | A | PAL_PStdBin | -1.92 | | A | | PAL_PStdBin | -2.44 |
| A | SHN_FxdBin | -0.22 | A | TRD_VarWgt | -0.54 | A | TRD_EquWgt | -1.32 | | A | | PAL_FxdBin | -2.32 |
| A | TRD_VarWgt | -0.01 | A | SHN_FxdBin | -0.44 | A | PAL_FxdBin | -1.18 | | A | | TRD_EquWgt | -1.73 |

| Lognormal Dist - Bias: None | | | Lognormal Dist - Bias: Pos 111 | | | Lognormal Dist - Bias: Pos 428 | | | Lognormal Dist - Bias: Pos 1500 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | Method | MAPE (%) | Pair-wise Signif | | | Method | MAPE (%) |
| C | SHN_Sheu | -33.75 | C | SHN_Sheu | -33.75 | C | SHN_Sheu | -33.83 | | D | | SHN_Sheu | -34.82 |
| B | SHN_StdBin | -14.70 | B | SHN_StdBin | -14.75 | B | SHN_StdBin | -15.00 | | | C | SHN_StdBin | -17.04 |
| A | SHN_PStdBin | -4.17 | A | SHN_PStdBin | -4.18 | A | TRD_VarWgt | -4.68 | | | C | TRD_VarWgt | -16.28 |
| A | SHN_OStdBin | -4.16 | A | SHN_OStdBin | -4.17 | A | SHN_PStdBin | -4.59 | B | | C | SHN_PStdBin | -9.88 |
| A | PAL_StdBin | -3.51 | A | PAL_StdBin | -3.55 | A | SHN_OStdBin | -4.55 | B | A | C | SHN_FxdBin | -8.80 |
| A | PAL_Sheu | -2.49 | A | PAL_Sheu | -2.53 | A | PAL_StdBin | -3.85 | B | A | C | SHN_OStdBin | -8.54 |
| A | ENT_GenMax | -2.17 | A | ENT_GenMax | -1.80 | A | PAL_Sheu | -2.80 | B | A | | PAL_StdBin | -6.59 |
| A | PAL_OStdBin | -1.73 | A | PAL_OStdBin | -1.75 | A | SHN_FxdBin | -2.40 | B | A | | PAL_FxdBin | -5.78 |
| A | PAL_PStdBin | -1.73 | A | PAL_PStdBin | -1.74 | A | PAL_OStdBin | -2.03 | B | A | | PAL_PStdBin | -5.72 |
| A | TRD_EquWgt | -1.19 | A | TRD_EquWgt | -1.21 | A | PAL_PStdBin | -2.03 | B | A | | PAL_Sheu | -5.55 |
| A | PAL_FxdBin | -0.79 | A | PAL_FxdBin | -0.83 | A | TRD_EquWgt | -1.48 | B | A | | PAL_OStdBin | -5.33 |
| A | SHN_FxdBin | -0.22 | A | TRD_VarWgt | -0.46 | A | PAL_FxdBin | -1.43 | B | A | | TRD_EquWgt | -4.66 |
| A | TRD_VarWgt | -0.01 | A | SHN_FxdBin | -0.40 | A | ENT_GenMax | -0.95 | | A | | ENT_GenMax | -0.37 |

1. When there is no bias under the assumption of 4 sources and 15 inputs from the lognormal distribution, SHN_Sheu's performance is not significantly different from PAL_Sheu in the case of estimating the population mean. However, for this configuration in the newsvendor problem, the PAL_Sheu is significantly better than SHN_Sheu.

2. For the case in which there is a negative bias of 300 or a positive bias of 428 or a negative bias with 8 sources and 5 inputs from a lognormal distribution, the TRD_VarWgt procedure is the second to worst performer. For this same configuration with the newsvendor problem, TRD_VarWgt is the second or third best performer.
3. For the configuration with 8 sources and 5 inputs from a lognormal population, ENT_GenMax is fourth from the worst performer when a negative bias of 300 or 600 is present. For the newsvendor problem with this same configuration, ENT_GenMax is the best performing data fusion method. The ENT_GenMax procedure under the no bias condition with 8 sources and 15 inputs from the lognormal distribution has a slightly better ranking for the newsvendor problem than for estimating the population mean.
4. For inputs from the lognormal distribution, SHN_Sheu is the best performer when 4 sources are used and the magnitude of the bias is large. This pattern does not carry over to the newsvendor problem. However, SHN_Sheu is still in the top performing grouping (not significantly different group) for the case with 4 sources.
5. For 15 inputs from a lognormal distribution, TRD_VarWgt is significantly the worst performing method in estimating the population mean with a negative bias of 600 or positive bias of 1500 when 8 sources are present. In this same configuration for the newsvendor problem, its performance improves somewhat.

## 4.6. Conclusions for Data Fusion Methods in the Newsvendor Model

The research questions for Essay 3 will be addressed in this section. This essay provides an analysis of the results for the newsvendor model that lead to both theoretical and practical recommendations of appropriate methods that are robust under various distributional and biased induced conditions. Thus, Essay 3 research questions contributes to the academic literature and practitioner guidance by providing insight into a systematic comparative study of the proposed data fusion methods in an SCM application, namely, the newsvendor model.

1. Do data fusion methods with relatively strong performance in estimating the parameter estimate also provide relatively strong performance in estimating the optimal demand under a given ratio of overage and underage costs?

In general, the top performers in Essay 2 for estimating the population mean repeat that performance in Essay 3. However, there are some inconsistencies. For example, SHN_Sheu is not performing at the same relative level as it did in estimating the mean of a population.

2. Do any of the data fusion methods deteriorate or improve on a relative basis with the introduction of positive and negative bias in estimating the optimal inventory level?

The data fusion method most affect by the type of bias is the ENT_GenMax, similar to the case when the data fusion methods were estimating the mean of a population.

3. Do the alternative entropy formulations to Shannon's entropy enhance the performance of the methods on a relative basis?

Many of Pal & Pal entropy formulations perform at least as their counterpart with Shannon entropy. This appears to be consistent for the performance of these data fusion methods in the newsvendor problem.

4. Is the relative rank order performance of the data fusion methods different in Essay 2 and Essay 3?

In general, the patterns are similar. However, several procedures have been shown to be quite different. As mentioned above, for the scenario having a negative bias of 300 or a positive bias of 428 with 8 sources and 5 inputs from a lognormal distribution, the TRD_VarWgt procedure is the second to worst performer in estimating the population mean. For this same configuration with the newsvendor problem, TRD_VarWgt is the second or third best performer.

## 4.7. Limitations to Results and Conclusions

This research focuses on 13 data fusion procedures, many are proposed modifications of Sheu (2010)'s data fusion methodology. The simulation study was conducted under a set of assumptions. These assumptions are mentioned in Chapter 1 and are repeated as follows:

1. Several sources provide "inputs" for information about descriptive parameters of data.
2. Heterogeneous distributions for the data from the sources make some sources less reliable.
3. Distortions, bias, censorship, and systemic errors may be more prominent in data from certain sources.
4. Sample size of data, the number of "inputs" from sources is generally small.

These assumptions may not be applicable to some data fusion applications. For example, the distributions may be homogeneous or the number of "inputs" may be very large. The conclusions of this study cannot extend to these situations although the simulation configurations were selected to represent or approximate a real world environment.

The bias was added to only the most reliable source to allow the bias to have the greatest impact since the most reliable source is usually weighted the most. Other distortions could have been added to data, such as error or noise to the values of inputs from a particular source. Data distortions such as censorship could have been included. To keep the number of simulation configurations manageable, only positive and negative bias were allowed as data distortions. Note that the lognormal scenario does not use symmetric bias values as the normal distribution scenario does. Since a lognormal is positively skewed, it is reasonable to believe that greater positive bias may be more likely to occur than negative bias. Also, the negative bias is limited as the values of the data are bounded below, but not above.

The distributions, the number of sources, and the sample sizes (“inputs”) were selected to represent situations in which data fusion methods may be of interest. In particular, time-sensitive data and small sample sizes from a few sources may present a challenge. This study could be extended to consider other distributions, such as the gamma or a “heavy tailed” distribution, to further examine the effects of the distribution on the data fusion methods.

Another contribution from this research is that a framework is explored to inform supply chain decision makers in selecting data fusion methods while knowing a method’s limitations from the simulation study. No data fusion method is completely robust to every challenge that a configuration can present. One recommendation is that a supply chain manager use several data fusion methods to assess differences and determine why differences may be occurring in the resulting estimations and applications.

REFERENCES


Alessandretti, G., Broggi, A., & Cerri, P. (2007). Vehicle and guard rail detection using radar and vision data fusion. *IEEE Transactions on Intelligent Transportation Systems, 8(1),* 95-105.

Allesina, S., Azzi, A., Battini, D., & Regattieri, A. (2010). Performance measurement in supply chains: new network analysis and entropic indexes. *International Journal of Production Research, 48(8),* 2297-2321.

Andersson, J., Jörnsten, K., Nonås, S. L., Sandal, L., & Ubøe, J. (2013). A maximum entropy approach to the newsvendor problem with partial information. *European journal of operational research, 228(1),* 190-200.

Arkhipov, A., & Ivanov, D. (2011). An entropy-based approach to simultaneous analysis of supply chain structural complexity and adaptation potential. *International Journal of Shipping and Transport Logistics, 3(2),* 180-197.

Bachmann, C., Abdulhai, B., Roorda, M. J., & Moshiri, B. (2013). A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling. *Transportation research part C: emerging technologies, 26*, 33-48.

Ballou, D., Madnick, S., & Wang, R. (2003). Special section: assuring information quality. *Journal of Management Information Systems*, 9-11.

Basir, O., & Shen, H. C. (1992). Sensory data integration: A team consensus approach. *1992 IEEE International Conference on Robotics and Automation*.

Bass, T. (1999). *Multisensor data fusion for next generation distributed intrusion detection systems.*

Bass, T. (2000). Intrusion detection systems and multisensor data fusion. *Communications of the ACM, 43(4)*, 99-105.

Bell, L. C., & Stukhart, G. (1986). Attributes of materials management systems. *Journal of Construction Engineering and Management, 112(1),* 14-21.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing, 58(3),* 239-258.

Bhandari, N., Koppelman, F. S., Schofer, J. L., Sethi, V., & Ivan, J. N. (1995). Arterial incident detection integrating data from multiple sources. *Transportation research record (1510),* 60-69.

Blair, W. D., Rice, T. R., Alouani, A. T., & Xia, P. (1991). Asynchronous data fusion for target tracking with a multitasking radar and optical sensor. Paper presented at the Orlando'91, Orlando, FL.

Byon, Y.-J., Shalaby, A., Abdulhai, B., & El-Tantawy, S. (2010). Traffic data fusion using SCAAT Kalman filters. *Paper presented at the Transportation Research Board 89th Annual Meeting.*

Carthel, C., Coraluppi, S., & Grignan, P. P. (2007). Multisensor tracking and fusion for maritime surveillance. *Paper presented at the FUSION.*

Chen, T., & Freeman, J. (2014). Using AHP-Entropy weight and TOPSIS methodology in green supplier selection. *Paper presented at the Proceedings of the European Operations Management Association Conference (EurOMA'14).*

Cheng, C.-H., Chang, J.-R., & Yeh, C.-A. (2006). Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost. *Technological Forecasting and Social Change, 73(5),* 524-542.

Cheu, R. L., Lee, D.-H., & Xie, C. (2001). An arterial speed estimation model fusing data from stationary and mobile sensors. *Paper presented at the Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE.*

Chung, A., Shen, H. C., & Basir, O. B. (1997). A decentralized approach to sensory data integration*. Paper presented at the Intelligent Robots and Systems, 1997. IEEE/RSJ International Conference on Robots and Systems*.

Cooper, M. C., & Ellram, L. M. (1993). Characteristics of supply chain management and the implications for purchasing and logistics strategy. *The international journal of logistics management, 4(2),* 13-24.

Cooper, M. C., Lambert, D. M., & Pagh, J. D. (1997). Supply chain management: more than a new name for logistics. *The international journal of logistics management, 8(1),* 1-14.

Chopra, S. & Meindl, P. (2013). *Supply chain management strategy, planning, and operation*. Pearson Publishers, Essex, England.

Coyle, J., Langley, C., Novack, R., & Gibson, B. (2012). *Supply chain management: a logistics perspective*: Cengage Learning

Curkovic, S. S., Thomas; Wagner Bret. (2015). Managing Supply Chain Risk: Integrating with Risk Management: *CRC Press*

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Intelligent Transportation Systems, 46(5*), 1416-1427.

Danese, P., Romano, P., & Formentini, M. (2013). The impact of supply chain integration on responsiveness: The moderating effect of using an international supplier network. *Transportation Research Part E* 49, 125-140.

Davis, T. (1993). Effective supply chain management. *Sloan management review*, 34, 35-35.

Dekkers, R., Kühnle, H., Amoo Durowoju, O., Kai Chan, H., & Wang, X. (2012). Entropy assessment of supply chain disruption. *Journal of Manufacturing Technology Management, 23(8),* 998-1014.

Dekkers, R., Kühnle, H., Gerschberger, M., Engelhardt-Nowitzki, C., Kummer, S., & Staberhofer, F. (2012). A model to determine complexity in supply networks. *Journal of Manufacturing Technology Management, 23(8),* 1015-1037.

Dell'Orco, M., & Teodorovic, D. (2009). Data fusion for updating information in modelling drivers' choice behaviour. *Transportmetrica, 5(2),* 107-123.

Diamond, S. M., & Ceruti, M. G. (2007). Application of wireless sensor network to military information integration. *Paper presented at the 2007 5th IEEE International Conference on Industrial Informatics.*

Djurdjanovic, D., Lee, J., & Ni, J. (2003). Watchdog Agent—an infotronics-based prognostics approach for product performance degradation assessment and prediction. *Advanced Engineering Informatics, 17(3),* 109-125.

Dong, X. L., & Naumann, F. (2009). Data fusion: resolving data conflicts for integration. *Proceedings of the VLDB Endowment, 2(2),* 1654-1655.

Durrant-Whyte, H., & Henderson, T. C. (2008). *Multisensor data fusion Springer Handbook of Robotics* (pp. 585-610): Springer

Easley, L. (2005). *System and method for fusion of container tracking data:* Google Patents.

Ehlers, M. (1991). Multisensor image fusion techniques in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 46(1),* 19-30.

El Faouzi, N.-E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges–A survey. *Information Fusion, 12(1),* 4-10.

English, L. P. (2009). *Information quality applied: Best practices for improving business information, processes and systems*: Wiley Publishing.

Eren, S., & Maglaras, C. (2006). Revenue management heuristics under limited market information: A maximum entropy approach. *Paper presented at the 6th Annual INFORMS Revenue Management Conference, June 2006.*

Fauvel, M., Chanussot, J., & Benediktsson, J. A. (2006). Decision fusion for the classification of urban remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 44(10)*, 2828-2838.

Fong, T., Thorpe, C., & Baur, C. (2001). Advanced interfaces for vehicle teleoperation: Collaborative control, sensor fusion displays, and remote driving tools. *Autonomous Robots, 11(1)*, 77-85.

Gan, H.-S., & Wirth*, A. (2005). Comparing deterministic, robust and online scheduling using entropy. *International Journal of Production Research, 43(10)*, 2113-2134.

Ghorbani, M., Arabzad, S. M., & Bahrami, M. (2012). Implementing Shannon entropy, SWOT and mathematical programming for supplier selection and order allocation. *International Journal of Supply Chain Management*, 1(1).

Ghorbani, M., Bahrami, M., & Arabzad, S. M. (2012). An integrated model for supplier selection and order allocation; using Shannon Entropy, SWOT and linear programming. *Procedia-Social and Behavioral Sciences, 41*, 521-527.

Girod, L., & Estrin, D. (2001). Robust range estimation using acoustic and multimodal sensing. Paper presented at the 2001 *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Goh, C., & Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management, 24(5)*, 511-517.

Goh, C., Law, R., & Mok, H. M. (2008). Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research, 46(3)*, 327-338.

Gong, Q., Jotshi, A., & Batta, R. (2004). Dispatching/routing of emergency vehicles in a disaster environment using data fusion concepts. *Paper presented at the Proceedings of 7th International Conference on Information Fusion*.

Grafström, A. (2010). Entropy of unequal probability sampling designs. *Statistical Methodology, 7(2)*, 84-97.

Greene, W. H. (2003). *Econometric analysis:* Pearson Education India.

Guoyi, X., & Xiaohua, C. (2011). Research on the third party logistics supplier selection evaluation based on AHP and entropy. *Paper presented at the 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC).*

Hall, D., & Llinas, J. (2001). *Multisensor data fusion*: CRC press.

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE, 85(1)*, 6-23.

Hall, D. L., & McMullen, S. A. (2004). *Mathematical techniques in multisensor data fusion*: Artech House.

Harrison, T. P., Lee, H. L., & Neale, J. J. (2005). *The practice of supply chain management: where theory and application converge*: Springer Science & Business Media.

Hendricks, K. B., & Singhal, V. R. (2005). Association between supply chain glitches and operating performance. *Management Science, 51*(5), 695-711.

Hess, A., & Fila, L. (2002). The joint strike fighter (JSF) PHM concept: potential impact on aging aircraft problems. *Paper presented at the Aerospace Conference Proceedings, 2002. IEEE.*

Hsu, C., & Wallace, W. A. (2007). An industrial network flow information integration model for supply chain management and intelligent transportation. *Information Systems, 1(3),* 327-351.

Hu, S., Zhu, X., Wang, H., & Koren, Y. (2008). Product variety and manufacturing complexity in assembly systems and supply chains. *CIRP Annals-Manufacturing Technology, 57(1),* 45-48.

Huatuco, L. H., Burgess, T. F., & Shaw, N. E. (2010). Entropic-related complexity for re-engineering a robust supply chain: a case study. *Production Planning & Control, 21(8),* 724-735.

Isik, F. (2010). An entropy-based approach for measuring complexity in supply chains. *International Journal of Production Research, 48(12),* 3681-3696.

Isik, F. (2011). Complexity in supply chains: A new approach to quantitative measurement of the supply-chain-complexity: *Intech Publisher*: Rijeka, Croatia.

Ivan, J. N. (1996). Real-time data fusion for arterial street incident detection using neural networks. *Transportation Research Part A, 1(30),* 60.

Ivan, J. N., Schofer, J. L., Koppelman, F. S., & Massone, L. L. (1995). Real-time data fusion for arterial street incident detection using neural networks. *Transportation research record (1497),* 27-35.

Jaber, M., Nuwayhid, R., & Rosen, M. (2006). A thermodynamic approach to modelling the economic order quantity. *Applied Mathematical Modelling, 30(9),* 867-883.

Jaber, M. Y. (2007). Lot sizing with permissible delay in payments and entropy cost. *Computers & Industrial Engineering, 52(1),* 78-88.

Jaber, M. Y., & Rosen, M. A. (2008). The economic order quantity repair and waste disposal model with entropy cost. *European Journal of Operational Research, 188(1),* 109-120.

Jaber, M. Y., Zanoni, S., & Zavanella, L. E. (2014). 'Consignment stock' for a two-level supply chain with entropy cost. *European Journal of Industrial Engineering, 8(2),* 244-272.

Janssens, D., Brijs, T., Vanhoof, K., & Wets, Geert (2006). Evaluating the performane of cost-based discretization versus entropy- and error-based discretization, *Computers & Operations Research* 33, 3107-3123.

Jiang, H., Zhao, S., Qiu, S., & Chen, Y. (2012). Strategy for technology standardization based on the theory of entropy. *Information Technology and Management, 13(4),* 311-320.

Jotshi, A., Gong, Q., & Batta, R. (2009). Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. *Socio-Economic Planning Sciences, 43(1),* 1-24.

Khaleghi, B., Khamis, A., & Karray, F. (2010). Random finite set theoretic based soft/hard data fusion with application for target tracking. *Paper presented at the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010*.

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information Fusion, 14(1),* 28-44.

Khan, S., Ganguly, A. R., & Gupta, A. (2008). Data mining and data fusion for enhanced decision support *Handbook on Decision Support Systems 1* (pp. 581-608): Springer.

Klein, L. A. (1993). *Sensor and data fusion concepts and applications.*

Kong, Q.-J., Li, Z., Chen, Y., & Liu, Y. (2009). An approach to urban traffic state estimation by fusing multisource information. *IEEE Transactions on Intelligent Transportation Systems, 10(3),* 499-511.

Kull, T. J. & Wacker, J. G (2010) Quality management effectiveness in Asia: The influence of culture. *Journal of Operations Management*, 28, 223-239.

Lan, Y., Gao, H., Ball, M. O., & Karaesmen, I. (2008). Revenue management with limited demand information. *Management Science, 54(9),* 1594-1609.

Lee, H. L., & Billington, C. (1992). Managing supply chain inventory: pitfalls and opportunities. *Sloan management review, 33(3).*

Lee, H. L., & Billington, C. (1995). The evolution of supply-chain-management models and practice at Hewlett-Packard. *Interfaces, 25(5),* 42-63.

Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006). Intelligent prognostics tools and e-maintenance. *Computers in industry, 57(6),* 476-489.

Lee, Y. W., & Strong, D. M. (2003). Knowing-why about data processes and data quality. *Journal of Management Information Systems, 20*(3), 13-39.

Lei, L., Liu, S., Ruszczynski, A., & Park, S. (2006). On the integrated production, inventory, and distribution routing problem. *IIE Transactions, 38(11),* 955-970.

Li, H., Manjunath, B., & Mitra, S. K. (1995). Multisensor image fusion using the wavelet transform. *Graphical models and image processing, 57(3),* 235-245.

Li, Y., Wang, L., & Heyde, M. (2010). Risk assessment of supply chain system based on information entropy. *Paper presented at the International Conference on Logistics Systems and Intelligent Management, 2010.*

Lim, A. E., & Shanthikumar, J. G. (2007). Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research, 55(2),* 198-214.

Liu, P., & Zhang, X. (2011). Research on the supplier selection of a supply chain based on entropy weight and improved ELECTRE-III method. *International Journal of Production Research, 49(3),* 637-646.

Llinas, J. (2002). Information fusion for natural and man-made disasters. *Paper presented at the Proceedings of the Fifth International Conference on Information Fusion, 2002.*

Lummus, R. R., & Vokurka, R. J. (1999). Defining supply chain management: a historical perspective and practical guidelines. Industrial Management & Data Systems, 99(1), 11-17.

Madnick, S., Wang, R., & Xian, X. (2003). The design and implementation of a corporate householding knowledge processor to improve data quality. *Journal of Management Information Systems, 20*(3), 41-70.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998) *Forecasting: methods and applications.* Wiley, New York.

Mangolini, M. (1994). *Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation.* Université de Nice Sophia-Antipolis.

Martínez-Olvera, C. (2008). Entropy as an assessment tool of supply chain information sharing. *European journal of operational research, 185(1),* 405-417.

McGill, J. I. (1995). Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints. *Annals of Operations Research, 60*(1), 209-240.

Metternicht, G., Hurni, L., & Gogu, R. (2005). Remote sensing of landslides: An analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote sensing of Environment, 98(2),* 284-303.

Mittelhammer, R., Cardell, N. S., & Marsh, T. L. (2013), The Data-Constrained Generalized Maximimum Entropy Estimator of the GLM: Asymptotic Theory and Inference. *Entropy*, 2013, 12, 1756-1775.

Nelson, P., & Palacharla, P. (1993). A neural network model for data fusion in ADVANCE. *Paper presented at the Pacific Rim TransTech Conference (1993: Seattle, Wash.). Proceedings Pacific Rim TransTech Conference. Vol. 1.*

Nichol, J., & Wong, M. (2005). Satellite remote sensing for detailed landslide inventories using change detection and image fusion. International journal of remote sensing, 26(9), 1913-1926.

Oliva, R., & Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and Operations Management, 18*(2), 138-151.

Opricovic, S. (1998). Multicriteria optimization of civil engineering systems. *Faculty of Civil Engineering, Belgrade, 2(1),* 5-21.

Pal, N. R., & Pal, S. K. (1991). Entropy: A new definition and its applications. *IEEE Transactions on Systems, Man and Cybernetics, 21(5)*, 1260-1270.

Perakis, G., & Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations Research, 56(1),* 188-203.

Pohl, C., & Van Genderen, J. L. (1998). Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International journal of remote sensing, 19(5)*, 823-854.

Pyle, D. (2003). *Business modeling and data mining:* Morgan Kaufmann.

Qiu, R. G. (2002). A data fusion framework for an integrated plant-wide information system. *Paper presented at the Proceedings of the Fifth International Conference on Information Fusion, 2002.*

Razavi, S. N., & Haas, C. T. (2010). Multisensor data fusion for on-site materials tracking in construction. *Automation in Construction, 19(8),* 1037-1046.

Redman, T. C. (1995). Opinion: improve data quality for competitive advantage. *Sloan management review, 36*(2), 99.

Roy, B. (1968). Classement et choix en présence de points de vue multiples. Revue française d'automatique, d'informatique et de recherche opérationnelle. *Recherche opérationnelle, 2(1),* 57-75.

Saaty, T. L. (2000). *Fundamentals of decision making and priority theory with the analytic hierarchy process (Vol. 6)*: Rws Publications.

Saghafian, S., & Tomlin, B. (2014). The newsvendor under demand ambiguity: Combining data with moment and tail information. *Available at SSRN 2388229*.

SAS Institute Inc. (2014). SAS/ETS® 13.2 *User's Guide: Cary, NC: SAS Institute Inc*.

Scholz-Reiter, B., Tervo, J. T., & Hinrichs, U. (2007*). Entropy as a measurement for the quality of demand forecasting Digital Enterprise Technology* (pp. 433-440): Springer.

Schulz, S. F., & Blecken, A. (2010). Horizontal cooperation in disaster relief logistics: Benefits and impediments. *International Journal of Physical Distribution & Logistics Management, 40*(8/9), 636-656. doi:10.1108/09600031011079300

Scott, P., & Rogova, G. (2004). Crisis management in a data fusion synthetic task environment. *Proceedings of FUSION 2004*.

Shannon, C. E. (1948). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5(1),* 3-55.

Shannon, C. E., & Weaver, W. (1963). *The mathematical theory of communication*: University of Illinois press.

Shapiro, B. P. (1977). *Sales Program Management: Formulation and Implementation:* McGraw-Hill Companies.

Shemshadi, A., Shirazi, H., Toreihi, M., & Tarokh, M. J. (2011). A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. *Expert Systems with Applications, 38(10),* 12160-12167.

Sheu, J.-B. (2007). An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transportation Research Part E: Logistics and Transportation Review, 43(6),* 687-709.

Sheu, J.-B. (2010). Dynamic relief-demand management for emergency logistics operations under large-scale disasters. *Transportation Research Part E: Logistics and Transportation Review, 46(1),* 1-17.

Shuiabi, E., Thomson, V., & Bhuiyan, N. (2005). Entropy as a measure of operational flexibility. *European journal of operational research, 165(3),* 696-707.

Siaterlis, C., & Maglaris, B. (2004). Towards multisensor data fusion for DoS detection. *Paper presented at the Proceedings of the 2004 ACM symposium on Applied Computing.*

Siraj, A., Vaughn, R. B., & Bridges, S. M. (2004). Intrusion sensor data fusion in an intelligent intrusion detection system architecture. *Paper presented at the Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.*

Sivadasan, S., Efstathiou, J., Frizelle, G., Shirazi, R., & Calinescu, A. (2002). An information-theoretic methodology for measuring the operational complexity of supplier-customer systems. *International Journal of Operations & Production Management, 22(1),* 80-102.

Smith, D., & Singh, S. (2006). Approaches to multisensor data fusion in target tracking: A survey. *IEEE Transactions on Knowledge and Data Engineering, 18(12),* 1696-1710.

Solberg, A. H. S., Jain, A. K., & Taxt, T. (1994). Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 32(4),* 768-778.

Song, J., Haas, C. T., Caldas, C., Ergen, E., & Akinci, B. (2006). Automating the task of tracking the delivery and receipt of fabricated pipe spools in industrial projects. *Automation in Construction, 15(2),* 166-177.

Southworth, F., Meyer, M., & Bronzini, M. (2008). *Defining Future Needs. Freight Demand Modeling: Tools for Public-Sector Decision Making,* 41-46.

Steinberg, A. N., Bowman, C. L., & White, F. E. (1999). Revisions to the JDL data fusion model. *Paper presented at the AeroSense 1999.*

Stroupe, A. W., Martin, M. C., & Balch, T. (2001). Distributed sensor fusion for object position estimation by multi-robot systems. *Paper presented at the Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation, 2001.*

Swaminathan, J. M., Smith, S. F., & Sadeh, N. M. (1998). Modeling supply chain dynamics: A multiagent approach. *Decision sciences, 29(3),* 607-632.

Tan, H.-S., Warf, G. K., & Henry, L. (2010). *Automated asset positioning for location and inventory tracking using multiple positioning techniques: Google Patents.*

Tarko, A., & Rouphail, N. (1993). Travel time data fusion in ADVANCE. *Paper presented at the Pacific Rim TransTech Conference (1993: Seattle, Wash.). Proceedings Pacific Rim TransTech Conference. Vol. 1.*

Thomas, D. J., & Griffin, P. M. (1996). Coordinated supply chain management. *European Journal of Operational Research, 94(1),* 1-15.

Treiber, M., Kesting, A., & Wilson, R. E. (2011). Reconstructing the traffic state by fusion of heterogeneous data. *Computer-Aided Civil and Infrastructure Engineering, 26(6),* 408-419.

Voelker, R. (2014). Ebola perspectives from opposite sides of the globe. *Jama, 312*(24). doi: 10.1001/jama.2014.13382

Wald, L. (1999). Definitions and terms of reference in data fusion. *Paper presented at the Joint EARSeL/ISPRS Workshop" fusion of sensor data, knowledge sources and algorithms for extraction and classification of topographic objects".*

Wang, H., Efstathiou, J., & Yang, J.-B. (2005). Entropy-based complexity measures for dynamic decision processes. *Dynamics of continuous discrete and impulsive systems series b, 12(5/6),* 829.

White Jr, F. E. (1987). Data fusion lexicon, joint directors of laboratories. *Technical panel for C, 3.*

Wu, M., & Liu, Z. (2011). The supplier selection application based on two methods: VIKOR algorithm with entropy method and Fuzzy TOPSIS with vague sets method. *International Journal of Management Science and Engineering Management, 6(2),* 109-115.

Xiu, G., & Chen, X. (2012). The third party logistics supplier selection and evaluation. *Journal of Software, 7(8),* 1783-1790.

Yoon, K. P., & Hwang, C.-L. (1995). *Multiple attribute decision making: an introduction (Vol. 104):* Sage publications.

Zhang, G., Shang, J., & Li, W. (2012). An information granulation entropy-based model for third-party logistics providers evaluation. *International Journal of Production Research, 50(1),* 177-190.

Zhang, J., & Xu, J. (2009). *Fuzzy Entropy Method for Quantifying Supply Chain Networks Complexity Complex Sciences* (pp. 1690-1700): Springer.

Zheng, X., Zhu, S., & Lin, Z. (2013). Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach. *Decision Support Systems,* 56, 211-222.