University of New Mexico
UNM Digital Repository

Mathematics & Statistics ETDs

Electronic Theses and Dissertations

1-28-2015

A Generalized Confidence Interval approach to comparing log-normal means, with application

Bose Falk

Follow this and additional works at: https://digitalrepository.unm.edu/math_etds

Recommended Citation

Falk, Bose. "A Generalized Confidence Interval approach to comparing log-normal means, with application." (2015). https://digitalrepository.unm.edu/math_etds/81

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Mathematics & Statistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Bose Falk

Candidate

Statistics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Guyoi Zhang, Chairperson

Yan Lu

Michael Sonksen

A Generalized Confidence Interval approach to comparing log-normal means, with application

by

Bose Falk

B.Sc., Economics and Mathematical Sciences
Open University, UK, 2011

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Statistics

The University of New Mexico

Albuquerque, New Mexico

December, 2014

©2014, Bose Falk

Dedication

I would like to dedicate this thesis to my wife Katerina, whose support, LaTeX editing tips and cups of tea were invaluable, as well as my baby daughter Tatiana for taking long enough naps to allow me to finish this thesis.

Acknowledgments

I would like to express my special appreciation and thanks to my advisor Professor Guoyi Zhang, who has helped tremendously with the research topic of the thesis, and my academic progress in general. I would also like to thank my committee members, professor Michael Sonksen and professor Yan Lu for serving as my committee members. I would also like to acknowledge my colleagues and contacts at 2degrees Network, for giving me the opportunity to learn about energy management in general, and the CRC scheme in particular.

A Generalized Confidence Interval approach to comparing log-normal means, with application

by

Bose Falk

B.Sc., Economics and Mathematical Sciences
Open University, UK, 2011

M.S., Statistics, University of New Mexico, 2014

Abstract

Generalized Confidence Intervals (GCI) can be constructed for cases where an exact confidence interval based on sufficient statistics is not available. In this thesis, we first review three existing tests for log-normal data using the GCI approach. Then we propose fiducial generalized pivotal quantities (FGPQ)-based simultaneous confidence intervals for ratios of log-normal means, and prove that the constructed confidence intervals have correct asymptotic coverage. These methods are then applied to a dataset from the Carbon Reduction Commitment Energy Efficiency Scheme (CRC) to test for differences between energy saving percentages among different groups.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Given a data set $\mathbf{x}$, for a certain significance level α and a parameter of interest θ, let $u(\mathbf{x})$ and $v(\mathbf{x})$ be functions of $\mathbf{x}$, such that $Pr(u(\mathbf{x}) \leq \theta \leq v(\mathbf{x})) = 1 - \alpha$, we call $(u(\mathbf{x}), v(\mathbf{x}))$ the $(1 - \alpha)$ confidence interval of θ. For most common statistical tests, confidence intervals can be constructed based on sufficient statistics using the central limit theorem.

However, for some distributions, it might not be possible to construct a confidence interval based on sufficient statistics. This is the case with the Exponential and the Log-normal distributions. For data that follows these distributions, for example, if we wish to do a comparison between two means, or to do inference for one mean, we can use the method of Generalized Confidence Intervals (GCI) (Weerahandi, 1993). In a GCI, the expressions for $u(\mathbf{x})$ and $v(\mathbf{x})$ include some random terms from known distributions, such as $Z \sim N(0, 1)$ and $U^2 \sim \chi^2_{n-1}$. The procedure continues with finding a generalized test statistic, T, for the specific distribution and θ we wish to test, then generates a large number of the random terms, and calculates a number of T_i from the distribution of T for $i = 1, \cdots, m$. The values for $u(\mathbf{x})$ and $v(\mathbf{x})$ are then given by using the appropriate percentiles of T.

The log-normal distribution is widely used to describe the distribution of positive random variables that exhibit skewness in biological, medical, economical and social studies. For example, the Carbon Reduction Commitment Energy Efficiency Scheme (CRC) is a mandatory energy usage reporting scheme for large organizations in the UK (refer to Section 3.1 for details). However, the percentage savings between two years for both the absolute and growth emissions variables do not follow a normal distribution even after transformation. After grouping the observations into representative groups, the variables are log-normally distributed. The problem of testing equality and multiple comparisons of the group means are common interests in many observational and experimental data arising from several populations. Unfortunately, if sample variances are unequal, the standard ANOVA tests don't apply for log-normal distributions even after transformation, since the null hypothesis based on log-transformed outcomes is not equivalent to the one based on the original outcomes (Zhou et al., 1997).

The problem of testing equivalence of the means of several log- normal populations has been well studied in literature. Approximation procedures are commonly used regarding this problem, for example, Alexander-Govern test (1994), the Welch test (1951) and the James second-order test (1951) etc. These three approximate tests behave similarly, and perform better than the ANOVA F-test (Guo & Luh, 2000). Later, Gupta and Li (2006) presented a score test. Weerahandi (1993), Krishnamoorthy and Mathew (2003) investigated inferences on the means of log-normal distributions by generalized p-values and GCIs. Bebu and Mathew (2008) proposed a GCI method for testing equivalence of bivariate response variables. Li (2009) proposed a new generalized p-value procedure.

Simultaneous confidence intervals for certain log-normal parameters are useful in many areas. For example, in order to investigate the effects of Automated Meter Reading (AMR) on energy savings, we can carry out a simultaneous comparison of the mean Absolute Emissions savings ratio of organizations with different levels of AMR (refer to Section 3.6 for details). To find out if there are AMR effects among different groups, we require a new method of multiple comparisons for several log-normal distributions. Hannig, Iyer, and Patterson (2006) introduced a subclass of Weerahandi's generalized pivotal quantities, called fiducial generalized pivotal quantities (FGPQs), and provided procedures to derive FGPQs. Hannig, Lidong, et al. (2006) proposed simultaneous fiducial GCI for ratios of means of log-normal distributions. Xiong and Mu (2009) proposed two kinds of simultaneous intervals based on FGPQ for all pairwise comparisons of treatment means in a one-way layout under heteroscedasticity. Xiong and Mu (2009) pointed out that if sample sizes are sufficiently large, Hannig, Lidong, et al. (2006)'s simultaneous confidence intervals are equal to one of their proposed intervals. Otherwise, Xiong and Mu (2009) methods perform better than Hannig, Lidong, et al. (2006)'s methods. Following Xiong and Mu (2009)'s idea, we proposed FGPQ-based simultaneous confidence intervals for all-pairwise comparisons for ratios of means from several log-normal populations under heteroscedasticity.

This thesis is outlined as follows. In Chapter 2, the methodology section, we review the log-normal distribution and discuss cases where it can be an appropriate modelling choice (Section 2.1). We then discuss the concepts behind the method of GCI as outlined by Weerahandi (1993) in Section 2.2. Then, we review three GCI test statistics for log-normal data: GCI for inference of one mean (Krishnamoorthy & Mathew, 2003) in Section 2.3, GCI for testing equivalence of bivariate response variables (Bebu & Mathew, 2008) in Section 2.4 and GCI for the difference between two independent log-normal means (Krishnamoorthy & Mathew, 2003) in Section 2.5. We then propose a FGPQ-based simultaneous confidence intervals for k log-normal means under heteroscedasticity and unbalanced design in Section 2.6. In Chapter 3, the analysis section, we apply the four tests to CRC data and have derived some interesting findings. In Chapter 4, the conclusions section, we summarize both theoretical and application results, and propose future work.

Chapter 2

Methodology

In this chapter, we first give an overview of the log-normal distribution and its possible applications. Next, we introduce the GCI concept from Weerahandi (1993). We then review three methods of comparing log-normal means using GCI: GCI for a log-normal mean (Krishnamoorthy & Mathew, 2003), GCI for equivalence of bivariate log-normal means (Bebu & Mathew, 2008) and GCI for two independent log-normal means (Krishnamoorthy & Mathew, 2003). Finally, we propose a new GCI for simultaneous pairwise comparisons of k log-normal means.

2.1 Log-Normal Distribution

Following Casella and Berger (2002, page 109), if the logarithm of a variable X is normally distributed, X follows the log-normal distribution - that is, $Y = log(X) \sim N(\mu, \sigma^2)$. The mean and variance of a log-normal distribution is given by:

$$E(X) = e^{(\mu+\frac{\sigma^2}{2})} \ ,$$

$$Var(X) = e^{(2\mu+\sigma^2)} - e^{2(\mu+\sigma^2)} \ ,$$

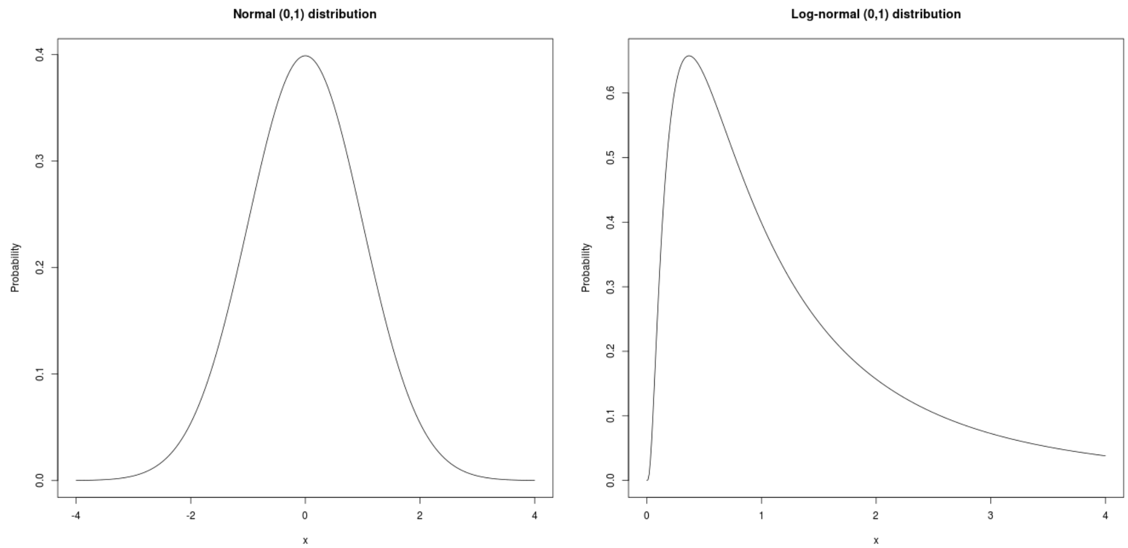

Figure 2.1: Probability density functions for a normal and log-normal distribution

where μ and σ^2 are the mean and variance of the normal distribution - see Figure 2.1 for a comparison of the probability density functions for a normal and log-normal distribution.

The log-normal distribution is widely used in a number of fields, for example ecology where the abundance of species can be modelled using a log-normal distribution (Magurran, 1998), geology where the concentration of elements in earth's crust are log-normally distributed (Malanca, 1996), in medicine where latency periods (the time between the infection and the first symptoms) are log-normal (Kondo, 1977), lingustics where the number of words per sentence are log-normal (Williams, 1940), network traffic (Antoniou et al., 2002), in economics where it can be used to model markets, for example incomes (Bundesamt fur Statistik, 1997), closing prices on stocks (Antoniou et al., 2004) and futures hedging (Lien & Balakrishnan, 2006).

In particular, the log-normal distribution is very useful for modelling populations which exhibit the following properties:

1. **Positive only values**: As the logarithm of negative numbers does not exist, it can only be used to model data which cannot take negative values, such as species population numbers and income levels. As a consequence, the log-normal distribution cannot take values < 0, and hence it can be a preferable model choice for cases where the normal distribution could give negative values but the data can never be negative.

2. **Increased skewness**: Some data generally follows a normal distribution for values close to the mean, but exhibit larger variance than expected at the tails. By using a log-normal distribution, this skewness is incorporated in the right-hand skew of the probability density function, and hence the log-normal distribution can be useful for modelling populations with larger than expected variances at the tail, such as market volatility.

3. **Multiplicative property**: The classical case of the Central Limit Theorem states that the sum of a sufficient number of independent random variables W_i is normally distributed (non-normalized version of Casella and Berger (2002, page 236)),

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W_i \to N\left(\mu, \frac{\sigma^2}{n}\right) \quad . \tag{2.1}$$

   If instead we multiply together W_i, $\prod_{i=1}^{n} W_i$ will tend towards a log-normal distribution as $log\left(\prod_{i=1}^{n} W_i\right) = \sum_{i=1}^{n} log(W_i)$.

To illustrate these three properties, we will use the example of stock market prices. It is commonly accepted that stock market prices are more appropriately modelled by a log-normal distribution than a normal distribution. To see why this is so, consider the following three properties of stock market prices (adapted from Daniel (2008) and Sharpe (2004)), which corresponds to the three properties outlined above:

1. **Positive values**: If the stock price followed a normal distribution, there is a possibility of the stock price taking a negative value, which is impossible. Stock prices can then only take positive values, fulfilling criteria 1 above.

2. **Market volatility**: There are many examples of market prices - whether in stocks, securities or cotton - showing much higher volatility (variance) than what would be expected were prices perfectly normally distributed. This results in fatter tails, and hence is an argument to use a log-normal distribution following the second criteria above.

3. **Rate of Return**: The rate of return on an investment, such as a stock, is the change in value of the stock, plus dividends paid, at the end of a certain time period such as one year. Using the notation from Daniel (2008) of initial stock price S_0, and the price after some interval k as S_k, the rate of return R_k for any one interval is given by

   $$R_k = \frac{S_k}{S_{k-1}} \ .$$

   Then $S_k = S_0(R_k * R_{k-1}...R_2 * R_1)$. If we assume that the R_ks are normally distributed and independent we see that unlike a normal CLT approximation which involves a sum of R_k, we are dealing with a product of R_ks, which approximates to a log-normal distribution as per the multiplicative property. Hence we can view as per Daniel (2008) the price of a stock at time t as a product of the initial price S_0 and a log-normal variable:

   $$S_t = S_0 e^{N(\mu t, \sigma^2 t)} = S_0 e^r \ .$$

   By dividing out the initial price S_0 from both sides of this equation we see the change in the dollar price s given by

   $$log(s) = R \ ,$$

which is the definition of a log-normal distribution. This exemplifies the multiplicative nature of the log-normal distribution, as each year's rate of change is a product of the previous year's rate of return - which corresponds to the third criteria above.

In the Analysis section, we will examine the CRC Energy Savings data and show that this also fulfills the three above properties, and hence assuming it passes the Shapiro-Wilks test on $Y = log(X)$ can be appropriately modeled using a log-normal distribution.

2.2 Generalized Confidence Intervals

The principles of GCI is outlined by Weerahandi (1993). The idea is to be able to construct confidence intervals for cases where we cannot construct exact confidence intervals based on sufficient statistics, such as for comparing two means from the exponential distribution, or the log-normal distribution. The confidence interval is constructed using a pivotal quantity (Weerahandi, 1993, page 900):

Let R be a function $r(\mathbf{X}; \mathbf{x}, \boldsymbol{\upsilon})$ where $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample, $\mathbf{x}$ are the observed values of $\mathbf{X}$, and $\boldsymbol{\upsilon} = (\theta, \boldsymbol{\delta})$ where θ is an unknown parameter of interest from $\mathbf{X}$ and $\boldsymbol{\delta}$ is a vector of nuisance parameters. Then R is a *generalized pivotal quantity* if it has the following two properties:

- Property A: R has a probability distribution free of unknown parameters
- Property B: The *observed pivotal*, defined as $r_{obs} = r(\mathbf{x}; \mathbf{x}, \boldsymbol{\upsilon})$ does not depend on the nuisance parameter $\boldsymbol{\delta}$.

Suppose that α is the desired significance level, such that the α-level confidence interval is given by $Pr(u(\mathbf{x}) \leq \theta \leq v(\mathbf{x})) = 1 - \alpha$. Then, given the pivotal quantity

R and the significance level α, we can define a subset of the sample space C_α:

$$Pr(R \in C(\mathbf{x})) = 1 - \alpha \ . \tag{2.2}$$

Define $\Theta_c(r)$ to be a subset of the parameter space such that:

$$\Theta_c(r) = \{\theta \in \Theta \ | r(\mathbf{x}; \mathbf{x}, \boldsymbol{\upsilon}) \in C_\alpha\} \ . \tag{2.3}$$

Following the argumentation in Theorem 2.1 of (Weerahandi, 1993), Θ_c is the generalized confidence region of θ.

Using this idea of generalized pivot quantities, one can explore the concept of a *generalized test variable* T. T is defined as $T = t(\mathbf{X}; \mathbf{x}, \boldsymbol{\upsilon})$, and has the same properties A and B as the generalized pivot quantity discussed above, as well as one additional:

- Property C: T is monotonically increasing in θ

Hence T is a general pivotal quantity with one additional restriction. We can then define the *generalized p-value* (Tsui & Weerahandi, 1989) as $p(t) = Pr(T \geq t | \theta = \theta_0)$, where t is the observed test variable, θ and θ_0 are given by the null hypothesis we wish to test:

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_\alpha : \theta > \theta_0 \ . \tag{2.4}$$

One can show that the function $\pi(T; \theta) = Pr(T \geq t | \theta)$ has a uniform(0,1) distribution, and hence we can use the generalized p-value $p(t)$ in the usual fashion, that is reject H_0 in Eq (2.4) when $p < \alpha$.

To summarize how to use this method in practice, once the equation for the generalized test statistic T has been found for a specific distribution and parameters to be tested, one can generate a number of simulated T_i, order these from $T_{min}, ..., T_{max}$

and the two-sided $(1-\alpha)$-level confidence interval is given by the $\alpha/2$ and $(1-\alpha/2)$ percentiles of t.

Using similar argumentation, outlined by Tsui and Weerahandi (1989), one can find a p-value for a hypothesis by looking at the mean of an indicator variable I_i, which takes the value $I_i = 1$ when t_i indicates that H_0 is true, and $I_i = 0$ when t_i indicates H_α is true. The generalized p-value is then given by $(1/m)\sum_{i=1}^{m} I_i$ (where $i = 1, ..., m$).

However, finding the pivotal quantities for a specific parameter for a specific distribution is non-trivial, and much work has been carried out to identify and test pivotal quantities. In Weerahandi (1993), the author outlines some methods for simplifying identification of pivotal quantities. Below, we will use three existing pivotal quantities / generalized test statistics for testing three different scenarios involving the log-normal distribution, which have been identified by the authors of the papers referenced in each sub-section. In addition we propose a new GCI test statistic for comparing several log-normal means. For each pivotal quantity, we use the method outlined here to generate suitable GCIs and generalized p-values for the specific problems we wish to address.

2.3 Confidence Interval for a Log-Normal Mean

As mentioned above, if the variable to be analyzed is log-normally distributed, that is: if X is a log-normally distributed variable, and μ and σ^2 are the mean and variance of $Y = log(X) \sim N(\mu, \sigma^2)$, the mean of the log-normal distribution is given by:

$$E(X) = e^{\eta} \quad \text{where} \quad \eta = \mu + \frac{\sigma^2}{2} \ .$$

If we wish to construct a confidence interval for η, using the method of GCI outlined above, we first need to identify the generalized test statistic T.

Krishnamoorthy and Mathew (2003) have identified such a test statistic T for drawing inferences on η. It is based on the following two sufficient statistics:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \; .$$

Let $\bar{y}$ and s^2 be the observed values of $\bar{Y}$ and S^2 respectively, the generalized test statistic T is a function of the random variables $\bar{Y}$, S^2, $\bar{y}$ and s^2 as follows:

$$T = \bar{y} - \frac{\bar{Y} - \mu}{S/\sqrt{n}} s/\sqrt{n} + \frac{1}{2}\frac{\sigma^2}{S^2} s^2 - \eta \; . \tag{2.5}$$

In order to remove the unknown values of μ and σ^2, we reduce them to known distributions using $Z = \sqrt{n}(\bar{Y} - \mu)/\sigma^2 \sim N(0,1)$, and independently $U^2 = (n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Substituting these into Eq (2.5) gives us:

$$T = \bar{y} - \frac{Z}{U/\sqrt{n-1}} \frac{s}{\sqrt{n}} + \frac{1}{2}\frac{s^2}{U^2/(n-1)} - \eta \; , \tag{2.6}$$

which satisfies the three required properties outlined above.

We are interested in the upper confidence interval for η only, hence we wish to test the null hypothesis:

$$H_0 : \eta \geq \eta_0 \quad \text{vs} \quad H_\alpha : \eta < \eta_0 \; . \tag{2.7}$$

For the upper confidence interval for η, we remove the term for η from Eq (2.6):

$$T = \bar{y} - \frac{Z}{U/\sqrt{n-1}} \frac{s}{\sqrt{n}} + \frac{1}{2}\frac{s^2}{U^2/(n-1)} \; , \tag{2.8}$$

where as above $Z \sim N(0,1)$ and $U^2 \sim \chi^2_{n-1}$. Hence, the generalized upper confidence interval of η can be derived by using the $(1-\alpha)$ percentile of T as defined in Eq (2.8).

The algorithm proposed by Krishnamoorthy and Mathew (2003) to construct this confidence interval is as follows:

Algorithm 1:

1. For a given data set $x_1, \ldots, x_n$, set $y_i = log(x_i)$, $i = 1, \ldots, n$

2. Compute $\bar{y} = (1/n)\Sigma_{i=1}^{n} y_i$ and $s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$

3. For $i = 1$ to m

   (a) Generate $Z \sim N(0,1)$ and $U^2 \sim \chi^2_{n-1}$

   (b) Set $T_i = \bar{y} - (Z/(U/\sqrt{n-1}))s/\sqrt{n} + \frac{1}{2}s^2/(U^2/(n-1))$

4. End i loop

5. Let $I_i = 1$ if $T_i \geq \eta_0$, else $I_i = 0$

6. The $100(1-\alpha)$ percentile of $T_1, \ldots, T_m$, denoted by $T(1-\alpha)$, is a Monte Carlo estimate of the $100(1-\alpha)$ generalized upper confidence interval for $\eta = \mu + \sigma^2/2$

7. $(1/m)\sum_{i=1}^{m} I_i$ is the generalized p-value for the null hypothesis in Eq (2.7)

We use the test statistic in Eq (2.8) and the algorithm above to construct a custom function in R to allow us to generate the generalized upper confidence interval for η. The function is referred to as *CIfunction(*η*,* η_0*,* m*,* α*)*, where η is the log-normal mean to be tested, η_0 is the constant in the null hypothesis, m is the number of iterations and α is the desired significance level. Details of how this function is set up can be found in Appendix B.

In simulations carried out by Krishnamoorthy and Mathew (2003), this method was shown to yield more accurate confidence intervals for small samples, than the parametric bootstrap method (Angus & Angust, 1994), or the large sample test method (Land, 1973) for constructing confidence interval for a log-normal mean.

2.4 Equivalence of Bivariate Log-Normal Means

In this section we identify a test for the equivalence of bivariate means from the log-normal distribution, based on the work of Bebu and Mathew (2008).

Let $(Y_1, Y_2)' = (log(X_1), log(X_2))'$ follow a bivariate normal distribution, where $\boldsymbol{\mu}$ is the means for $(Y_1, Y_2)'$ and the covariance matrix is denoted by Σ:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad , \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} .$$

By the definition of the means and variances of the log-normal distribution we have:

$$E(X_1) = e^{(\mu_1 + \frac{1}{2}\sigma_{11})} = e^{\eta_1} \quad \& \quad E(X_2) = e^{(\mu_2 + \frac{1}{2}\sigma_{22})} = e^{\eta_2} .$$

Let $\theta = \eta_1 - \eta_2 = (\mu_1 - \mu_2) + \frac{1}{2}(\sigma_{11} - \sigma_{22})$. To test

$$H_0 : \eta_1 = \eta_2 \quad \text{vs} \quad H_\alpha : \eta_1 \neq \eta_2 , \tag{2.9}$$

we wish to construct a confidence interval for θ.

If this confidence interval for θ includes 0 at some significance level α then we cannot reject the null hypothesis that η_1 and η_2 are equivalent, in other words that the two bivariate response variables are equivalent.

In order to construct a GCI for θ, we need to identify the generalized test statistic T. Bebu and Mathew (2008) discussed GCI for this θ. Let

$$\mathbf{A} = \sum_{i=1}^{n} \begin{pmatrix} Y_{1i} - \bar{Y}_1 \\ Y_{2i} - \bar{Y}_2 \end{pmatrix} \begin{pmatrix} Y_{1i} - \bar{Y}_1 \\ Y_{2i} - \bar{Y}_2 \end{pmatrix}' = \begin{pmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{pmatrix} .$$

Then, $(\bar{Y}_1, \bar{Y}_2)' \sim N(\boldsymbol{\mu}, (1/n)\Sigma)$ and $\mathbf{A} \sim W_2(\Sigma, n-1)$, where W_2 is the bivariate Wishart distribution. Using the properties of the Wishart distribution we define:

$$A_{11.2} = A_{11} - \frac{A_{12}^2}{A_{22}} \quad \text{and} \quad \sigma_{11.2} = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} ,$$

$$U_{22} = \frac{A_{22}}{\sigma_{22}} \sim \chi^2_{n-1} \ ,$$

$$U_{11.2} = \frac{A_{11.2}}{\sigma_{11.2}} \sim \chi^2_{n-2} \ ,$$

$$Z_1 = \frac{\left(A_{12} - \frac{\sigma_{12}}{\sigma_{22}} A_{22}\right)}{\sqrt{\sigma_{11.2} A_{22}}} \sim N(0,1) \ .$$

Using a_{ij} for the observed values of A_{ij}, we can define a matrix $\mathbf{R}$:

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{pmatrix} \ ,$$

$$R_{22} = \frac{a_{22}}{U_{22}} \ ,$$

$$R_{12} = \frac{a_{12}}{U_{12}} - \left(\sqrt{a_{11.2} a_{22}} \frac{Z_1}{\sqrt{U_{11.2}}} \frac{1}{U_{22}} \right) \ ,$$

$$R_{11} = \frac{a_{11.2}}{U_{11.2}} + \frac{R_{12}^2}{R_{22}} \ .$$

Then we can use these values to construct the generalized test statistic T:

$$T = (\bar{y}_1 - \bar{y}_2) - \frac{Z_2}{\sqrt{n}} \sqrt{R_{11} - 2R_{12} + R_{22}} + \frac{1}{2}(R_{11} - R_{22}) \ . \tag{2.10}$$

This test statistic T fulfills the three required properties outlined in Section 2.2, and can hence be used to construct the GCI for θ.

To generate the confidence interval we will generate a number of values T_i, where $i = 1, 2...m$ by randomly generating values for $U_{22} \sim \chi^2_{n-1}$, $U_{11.2} \sim \chi^2_{n-2}$ and $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$ and using them in Eq (2.10). The algorithm proposed by Bebu and Mathew (2008) to generate this confidence interval is as follows:

Algorithm 2:

1. For a given sample (x_{1i}, x_{2i}) from the log-normal distribution, set $(y_{1i}, y_{2i})' = (log(x_{1i}), log(x_{2i}))'$

2. Compute the sample mean $(\bar{y}_1, \bar{y}_2)$, and the matrix

$$\mathbf{a} = \sum_{i=1}^{n} \begin{pmatrix} y_{1i} - \bar{y}_1 \\ y_{2i} - \bar{y}_2 \end{pmatrix} \begin{pmatrix} y_{1i} - \bar{y}_1 \\ y_{2i} - \bar{y}_2 \end{pmatrix}' = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

3. For $j = 1$ to m

   (a) Generate $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$, $U_{22} \sim \chi^2_{n-1}$ and $U_{11.2} \sim \chi^2_{n-2}$

   (b) Compute R_{22}, $R_{11.2}$ and R_{11}

   (c) Compute

$$T = (\bar{y}_1 - \bar{y}_2) - \frac{Z_2}{\sqrt{n}} \sqrt{R_{11} - 2R_{12} + R_{22}} + \frac{1}{2}(R_{11} - R_{22})$$

4. End j loop

5. The $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of T provide a Monte Carlo estimate of the two-sided $100(1-\alpha)$ confidence interval for $\theta = (\mu_1 - \mu_2) + \frac{1}{2}(\sigma_{11} - \sigma_{22})$

The confidence interval for θ is then given by the $T_j(1 - \frac{\alpha}{2})$ and $T_j(\frac{\alpha}{2})$ quantiles of T_j. If this confidence interval includes 0, we cannot reject the null hypothesis that $\eta_1 = \eta_2$ at the α confidence level.

We use this method to investigate the equivalence of the Absolute and Growth response variables. We construct a custom function *bivariate(c, d, m, α)* in R, where we take the two variables $c = X_1$ and $d = X_2$, m is the number of iterations and α is the required significance level. See Appendix B for details of how this function is constructed.

2.5 Confidence Interval of Two Independent Log-Normal Means

The methodology for comparing the equivalence of two independent log-normally distributed means follow much the same structure as that for constructing a confidence interval for one mean outlined above, also from Krishnamoorthy and Mathew (2003). Let $Y_j = log(X_j) \sim N(\mu_j, \sigma_j^2)$, $j = A, B$. To compare the log-normal means from two samples A and B, η_j $(j = A, B)$, we wish to test the following hypothesis:

$$H_0 : \eta_A \geq \eta_B \quad \text{vs} \quad H_\alpha : \eta_A < \eta_B \ . \tag{2.11}$$

We need to find a generalized test statistic T which fulfills the three criteria of being free from unknown parameters, the observed value t being free of nuisance parameters, and T is stochastically increasing in θ. From Krishnamoorthy and Mathew (2003), such a test statistic is:

$$T_j = \bar{y}_j - \frac{\bar{Y}_j - \mu_j}{S_j/\sqrt{n_j}} \frac{s_j}{\sqrt{n_j}} + \frac{1}{2} \frac{\sigma_j^2}{S_j^2} s_j^2 \ , \tag{2.12}$$

where $j = A, B$. As before, $\bar{y}_j$ is the observed mean of Y_j, s_j^2 is the observed variance of Y_j, and n_j is the sample size for group j. We can then replace $\bar{Y}_j$, μ_j, σ_j^2 and S_j^2 with $Z_j \sim N(0, 1)$ and $U_j^2 \sim \chi^2_{n_j - 1}$ to obtain:

$$T_j = \bar{y}_j - \frac{Z_j}{U_j/\sqrt{n_j - 1}} \frac{s_j}{\sqrt{n_j}} + \frac{1}{2} \frac{s_j^2}{U_j^2/(n - 1_j)} \ . \tag{2.13}$$

Using the definition of T_j as per Eq (2.13), we can calculate the test statistic for the null hypothesis in Eq (2.11) as

$$T_2 = T_A - T_B \ , \tag{2.14}$$

which allows us to calculate the GCI for T_2. This is again done by generating m random values for $Z_j \sim N(0, 1)$ and $U_j^2 \sim \chi^2_{n-1}$, and finding the proportion $T_{2i} \leq 0$

(where $i = 1, 2...m$). The two-sided confidence interval for T_2, the differences between the log-normal means of group A and group B, is again the $(1 - \frac{\alpha}{2})$ quantile and $\frac{\alpha}{2}$ quantile of of T_2 respectively.

The algorithm to construct this confidence interval is the same as Algorithm 1 per Section 2.3 to generate T_A and T_B, and the test statistic T_2 is calculated using Eq (2.14).

This method is used for finding whether the savings are significant for organizations in two subpopulations, such as those where $AMR = 0$ vs $AMR > 0$. In R, we construct a custom function *comparison(a, b, m, α)* which will in the notation above take as inputs the data sets A and B, the number of iterations m and the significance level α, and output a p-value for the hypothesis in Eq (2.11) and a confidence interval for $\eta_A - \eta_B$. Details of this function can be found in Appendix B.

This test was compared by Krishnamoorthy and Mathew (2003) to the large sample test for one- and two- sided large sample Z-score (Zhou & Gao, 1997), and it was shown that the method outlined above produced more accurate confidence intervals for small samples sizes.

2.6 Inference of Several Log-Normal Distributions

In this section we propose FGPQ-based simultaneous confidence intervals for comparing several log-normal means when variances are heteroscedastic and group sizes are unequal. We also prove that the constructed confidence intervals have correct asymptotic coverage. Simulation studies show that the proposed confidence intervals work well even for small samples.

2.6.1 Notation and Li's overall test

First we briefly review and define notation. Let X_{ij}, where $i = 1, ..., k$ and $j = 1, ..., n_i$ be random samples from k log-normal distributions, with parameters μ_i and σ_i^2, and let $Y_{ij} = log(X_{ij})$. Let $\bar{Y}_i$ and S_i^2 be the observed mean and variance of Y_i, and as per Section 2.4 define

$$Z_i = \sqrt{n_i}(\bar{Y}_i - \mu_i)/\sigma_i \sim N(0,1) \ ,$$

$$U^2 = (n_i - 1)S_i^2/\sigma_i^2 \sim \chi^2_{n_i-1} \ .$$

Then, from the expectations of the log-normal distribution we define

$$M_i = E(X_{ij}) = e^{\mu_i + \sigma_i^2/2} \quad \text{and} \quad \theta_i = log(M_i) = \mu_i + \sigma_i^2/2 \ . \tag{2.15}$$

As discussed in section 2.3 and 2.5, from Krishnamoorthy and Mathew (2003), we have the following generalized pivot statistics:

$$T_{\mu_i} = \bar{y}_i - \sqrt{\frac{n_i - 1}{n_i}} \cdot \frac{Z_i s_i}{U_i} \quad \text{and} \quad T_{\sigma_i^2} = \frac{s_i^2}{U_i^2/(n_i - 1)} \ . \tag{2.16}$$

In the following, we will also review the overall tests of several log-normal groups proposed by Li (2009). The hypothesis of interest is:

$$H_0 : M_1 = M_2 = \ldots = M_k \quad \text{vs} \quad H_\alpha : \text{ Not all } M_i \text{ are equal} \ ,$$

which is equivalent to

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \quad \text{vs} \quad H_\alpha : \text{ Not all } \theta_i \text{ are equal} \ . \tag{2.17}$$

Let $\mathbf{H}$ be a matrix with dimensions $(k-1) \times k$:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \ ,$$

and let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)'$, $\bar{\mathbf{Y}} = (\bar{Y}_1, \bar{Y}_2, \cdots, \bar{Y}_k), \mathbf{S}^2 = (S_1^2, S_2^2, \cdots, S_k^2)$, $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_k)$, and $\mathbf{s}^2 = (s_1^2, s_2^2, \cdots, s_k^2)$.

Then, using the generalized test statistic from Eq (2.16), we find the pivotal quantities for θ_i and $\mathbf{H}\boldsymbol{\theta}$ to be:

$$\theta_i = T_{\mu_i} + T_{\sigma_i^2}/2$$

$$T_{\mathbf{H}\boldsymbol{\theta}} = \mathbf{H}(T_{\theta_1}, \cdots, T_{\theta_k})'.$$

The pivotal quantity is then $T_{\mathbf{H}\boldsymbol{\theta}}$, whose mean and variance are given by:

$$\mu_T = E(T_{\mathbf{H}\boldsymbol{\theta}}|(\bar{\mathbf{y}}, \mathbf{s}^2)) = \mathbf{H}E(T_{\boldsymbol{\theta}}|(\bar{\mathbf{y}}, \mathbf{s}^2))$$

$$\boldsymbol{\Sigma}_T = \text{cov}(T_{\mathbf{H}\boldsymbol{\theta}}|(\bar{\mathbf{y}}, \mathbf{s}^2)) = \mathbf{H}\text{cov}(T_{\boldsymbol{\theta}}|(\bar{\mathbf{y}}, \mathbf{s}^2))\mathbf{H}' \ .$$

Then, the generalized p-value is given by

$$p = P\left\{(T_{\mathbf{H}\boldsymbol{\theta}} - \mu_T)'\boldsymbol{\Sigma}_T^{-1}(T_{\mathbf{H}\boldsymbol{\theta}} - \mu_T) \geq \mu_T'\boldsymbol{\Sigma}_T^{-1}\mu_T\right\}. \tag{2.18}$$

To calculate the generalized p-values using this method, Li (2009) suggested the following algorithm:

Algorithm 3:

1. For a given sample with group sizes $(n_1, \ldots, n_k)$, group means $(\bar{y}_1, \ldots, \bar{y}_k)$ and variances $(s_1^2, \ldots, s_k^2)$
2. For $l = 1$ to L
   (a) Generate Z_i and U_i, where $i = 1, \ldots, k$
   (b) Compute $T_l = T_{\mathbf{H}\boldsymbol{\theta}} = \mathbf{H}(T_{\theta 1}, \ldots, T_{\theta k})'$
   (c) Compute $\hat{\mu}_T = \frac{1}{L}\sum_{l=1}^{L} T_l$ and $\hat{\Sigma}_T = \frac{1}{L-1}\sum_{l=1}^{L}(T_l - \hat{\mu}_T)(T_l - \hat{\mu}_T)'$

(d) Compute $\|\tilde{\hat{T}}\|_l^2 = (T_l - \hat{\mu}_T)'\hat{\Sigma}_T^{-1}(T_l - \hat{\mu}_T)$, where $l = 1, \ldots, L$

(e) Compute $\|\tilde{\hat{\mu}}_0\|^2 = \hat{\mu}_T'\hat{\Sigma}_T^{-1}\hat{\mu}_T$

(f) Let $W_l = 1$ if $\|\tilde{\hat{T}}\|_l^2 \geq \|\tilde{\hat{\mu}}_0\|^2$, otherwise $W_l = 0$

3. End l loop

4. The generalized p-value for Eq (2.17) is given by $(1/L)\sum_{l=1}^{L} W_l$

2.6.2 GCI for simultaneous pairwise comparison

In order to construct a GCI for simultaneous pairwise comparisons of k log-normal means, we first need to define the test we wish to carry out:

$$H_0 : M_i = M_j \text{ for all } i \neq j \quad \text{vs} \quad H_\alpha : \text{ at least one } M_i \neq M_j \ .$$

If we define the ratio of means as $M_{ij} = M_i/M_j$, and

$$\theta_{ij} = log(M_{ij}) = log\left(\frac{M_i}{M_j}\right) = log\left(\frac{e^{\mu_i+\sigma_i^2/2}}{e^{\mu_j+\sigma_j^2/2}}\right) = \left(\mu_i - \frac{\sigma_i^2}{2}\right) - \left(\mu_j - \frac{\sigma_j^2}{2}\right) \ ,$$

then, since constructing a confidence interval for $M_{ij} = M_i/M_j$ is equivalent to constructing one for θ_{ij} as defined above, the null hypothesis we wish to test can be expressed as

$$H_0 : \text{All } \theta_{ij} = 0 \quad \text{vs} \quad H_\alpha : \text{Not all } \theta_{ij} = 0 \ .$$

Following Xiong and Mu (2009), we can define the FGPQs for μ_i and σ_i^2 (for $i = 1, ..., k$) as R_{μ_i} and $R_{\sigma_i^2}$:

$$R_{\mu_i} = \bar{Y}_i - \sqrt{\frac{n_1 - 1}{n_i}} \cdot \frac{S_i Z_i}{U_i} \ ,$$

$$R_{\sigma_i^2} = \frac{(n_i - 1)S_i^2}{U_i^2} \ ,$$

which then gives us the pivotal statistic for θ_i, R_{θ_i}:

$$R_{\theta_i} = R_{\mu_i} - \frac{R_{\sigma_i^2}}{2} = \bar{Y}_i - \sqrt{\frac{n_i-1}{n_i}} \cdot \frac{S_i^2 Z_i}{U_i} + \frac{(n_i-1)S_i^2}{2U_i^2} \ .$$

Using this defintion of R_{θ_i}, we can define $R_{\theta_{ij}}$ as

$$R_{\theta ij} = \bar{Y}_i - \bar{Y}_j - \sqrt{\frac{n_i-1}{n_i}}\frac{S_i Z_i}{U_i} + \sqrt{\frac{n_j-1}{n_j}}\frac{S_j Z_j}{U_j} + \frac{(n_i-1)S_i^2}{2U_i^2} - \frac{(n_j-1)S_j^2}{2U_j^2} \ . \tag{2.19}$$

Then, the expectation η and variance V of $R_{\theta_{ij}}$, conditional on the observed values of $\bar{y}$ and s^2 for groups i and j are:

$$\eta_{ij} = E(R_{\theta_{ij}}|\bar{\mathbf{Y}}, \mathbf{S}^2) = \bar{Y}_i - \bar{Y}_j + \frac{n_i-1}{2(n_i-3)}S_i^2 - \frac{n_i-1}{2(n_i-3)}S_i^2 \ , \tag{2.20}$$

$$\begin{aligned} V_{ij} = Var(R_{\theta_{ij}}|\bar{\mathbf{Y}}, \mathbf{S}^2) = & \frac{n_i-1}{n_i(n_i-3)}S_i^2 + \frac{(n_i-1)^2}{2(n_i-3)^2(n_i-5)}S_i^4 \\ & + \frac{n_j-1}{n_j(n_j-3)}S_j^2 + \frac{(n_j-1)^2}{2(n_j-3)^2(n_j-5)}S_j^4 \ . \end{aligned} \tag{2.21}$$

If ξ_{ij} is the variance of η_{ij}, and $R_{\xi ij}$ the pivotal statistic of ξ_{ij}, then:

$$\begin{aligned} \xi_{ij} = & \ \mathrm{Var}\{E(R_{\theta_{ij}}|\bar{\mathbf{Y}}, \mathbf{S}^2)\} \\ = & \frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j} + \left(\frac{n_i-1}{2(n_i-3)}\right)^2 \frac{2\sigma_i^4}{n_i} + \left(\frac{n_j-1}{2(n_j-3)}\right)^2 \frac{2\sigma_j^4}{n_j} \\ = & \frac{\sigma_i^2}{n_i} + \frac{(n_i-1)^2}{2n_i(n_i-3)^2} + \frac{\sigma_j^2}{n_j} + \frac{(n_j-1)^2}{2n_j(n_j-3)^2} \ , \end{aligned} \tag{2.22}$$

$$\begin{aligned} R_{\xi ij} = & \frac{(n_i-1)S_i^2}{n_i U_i^2} + \frac{(n_i-1)^2}{2n_i(n_i-3)^2}\left(\frac{(n_i-1)S_i^2}{U_i^2}\right)^2 \\ & + \frac{(n_j-1)S_j^2}{n_j U_j^2} + \frac{(n_j-1)^2}{2n_j(n_j-3)^2}\left(\frac{(n_j-1)S_j^2}{U_j^2}\right)^2 \ . \end{aligned} \tag{2.23}$$

Using FGPQs to approximate distributions as per (Xiong & Mu, 2009), we can approximate the distribution

$$\max_{i \leq j} \left| \frac{\theta_{ij} - E(R_{\theta_{ij}} | \bar{\mathbf{Y}}, \mathbf{S}^2)}{\sqrt{\mathrm{Var}(R_{\theta_{ij}} | \bar{\mathbf{Y}}, \mathbf{S}^2)}} \right| , \tag{2.24}$$

with

$$Q = \max_{i \leq j} \left| \frac{R_{\theta_{ij}} - E(R_{\theta_{ij}} | \bar{\mathbf{Y}}, \mathbf{S}^2)}{\sqrt{R_{\epsilon_{ij}}}} \right| . \tag{2.25}$$

Then, the simultaneous confidence intervals for θ_{ij} is

$$\eta_{ij} \pm q(\alpha)\sqrt{V_{ij}} , \tag{2.26}$$

where $q(\alpha)$ is the upper α quantile of Q.

Based on the following theorem, the confidence interval in Eq (2.26), have the correct coverage probabilities - the proof of this theorem is available in Appendix C.

Theorem 1. *Let $Y_{i1}, \cdots, X_{in_i}, i = 1, \cdots, k$ be random samples from k different populations and be mutually independent. Assume that $0 < \sigma_i^2 = Var(Y_{i1}) < \infty, \mu_i = E(Y_{i1}), N = \sum_{i=1}^{k} n_i$ and $\frac{n_i}{N} \to \lambda_i \in (0,1)$ as $N \to \infty$ for all i, then*

$$P(\theta_{ij} \in \eta_{ij} \pm q(\alpha)\sqrt{V_{ij}} \quad \textit{for all} \quad i < j) \xrightarrow{p} 1 - \alpha.$$

The algorithm proposed to use the above method to generate confidence intervals is as follows:

Algorithm 4:

1. For given observations x_{ij} (where $i = 1, \ldots, k,\ j = 1, \ldots, n_i$) compute $y_{ij} = log(x_{ij})$

2. Compute $\bar{y}_i$ and s_i^2 (where $i = 1, \ldots, k$)

3. For $l = 1, \ldots, m$

   (a) Generate $Z_i \sim N(0,1)$ and $U_i^2 \sim \chi^2_{n_i-1}$

   (b) Compute $R_{\theta_{ij}}$, $R_{\xi_{ij}}$ and Q_l

4. End l loop

5. The confidence interval for θ is then given by computing $q(\alpha)$, the $100(1-\alpha)$ percentile of Q

In R, we create a function *severalmeans(data, m, α)*, where *data* is an array with the k groups we wish to compare, m is the number of iterations for each comparison, and α is the desired significance level. The output is the two-sided confidence interval for each $\theta_i - \theta_j$, and if this interval includes 0 we cannot reject the null hypothesis that the two log-normal means are the same. Details of how this function is calculated are available in Appendix B.

2.6.3 Simulations

To ensure the above methodology gives the correct coverage probabilities, simulations were carried out, following similar settings as per Li (2009). $\bar{y}_i$ and s_i^2 are generated using $\bar{y}_i \sim N(0, \sigma^2/n_i)$ and $s_i^2 \sim \sigma_i^2 \chi^2_{n_i-1}/(n_i-1)$, where $0 \leq \sigma^2 \leq 1$, and $i = 2, ..., k$.

The simulations has the following parameters:

1. Number of groups k: $k = 3$ and $k = 6$

2. Population variance $\boldsymbol{\sigma} = (\sigma_1^2, ..., \sigma_k^2)$: various combinations

3. Population mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)$: various combinations

4. Significance level α: $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.1$

5. Group sizes $\mathbf{n} = (n_1, ..., n_k)$: various combinations

For a given sample size and parameter configuration, we generated 2000 observed vectors $(s_1^2, ..., s_k^2, \bar{y}_1, ..., \bar{y}_k)$ and used 5000 runs to estimate the Type 1 errors (simulated p-value). According to our experience, 5000 runs is sufficient to guarantee the precision of simulated p-value. We consider both overall tests (Li, 2009) and multiple comparisons (proposed method). Algorithm 3 is used to estimate the simulated p-value of overall test. Algorithm 4 is used to find q_α, the $1 - \alpha$ percentile of the simulated distribution of Q.

Tables 1 and 2 report the simulation results of Li (2009)'s overall test and the proposed multiple comparison procedure under different settings. We can see that simulated p-values of the overall test and multiple comparison procedure are close to the nominal levels when the group sizes are 10 or more. We found that the overall tests and the proposed MCP perform well for both unbalanced unequal variance and balanced equal variance cases.

Table 2.1: Simulation result 1

$\mathbf{n} = (n_1, n_2, n_3)$ is a vector of unequal group sizes with $\mathbf{n}^{(1)} = (10, 16, 20), \mathbf{n}^{(2)} = (10, 10, 10), \mathbf{n}^{(3)} = (20, 16, 10)$; $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ is a vector of unequal means, with $\boldsymbol{\mu}^{(1)} = (1, 1, 1), \boldsymbol{\mu}^{(2)} = (1, 1, 1.25), \boldsymbol{\mu}^{(3)} = (1, 1.25, 1.45)$; $\boldsymbol{\sigma} = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ is a vector of unequal variances, with $\boldsymbol{\sigma}^{(1)} = (0.1, 0.1, 0.1), \boldsymbol{\sigma} = (1, 1, 0.5), \boldsymbol{\sigma}^{(3)} = (1, 0.5, 0.1)$. "Overall" means overall test from Li, 2009 for equality of group means; "MCP" is the FGPQ-based multiple comparison procedure (proposed method); Numbers in Table are simulated p-values.

		$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
$\mathbf{n}$	$(\boldsymbol{\mu}, \boldsymbol{\sigma})$	Overall	MCP	Overall	MCP	Overall	MCP
$\mathbf{n}^{(1)}$	$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\sigma}^{(1)})$	0.0070	0.0105	0.0405	0.0475	0.0800	0.0840
	$(\boldsymbol{\mu}^{(2)}, \boldsymbol{\sigma}^{(2)})$	0.0065	0.0150	0.0395	0.0410	0.0835	0.0825
	$(\boldsymbol{\mu}^{(3)}, \boldsymbol{\sigma}^{(3)})$	0.0060	0.0170	0.0390	0.0455	0.0845	0.0800
$\mathbf{n}^{(2)}$	$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\sigma}^{(1)})$	0.0070	0.0150	0.0385	0.0410	0.0735	0.0765
	$(\boldsymbol{\mu}^{(2)}, \boldsymbol{\sigma}^{(2)})$	0.0060	0.0105	0.0410	0.0450	0.0870	0.0830
	$(\boldsymbol{\mu}^{(3)}, \boldsymbol{\sigma}^{(3)})$	0.0050	0.0140	0.0435	0.0385	0.0790	0.0830
$\mathbf{n}^{(3)}$	$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\sigma}^{(1)})$	0.0070	0.0100	0.0445	0.0400	0.0785	0.0890
	$(\boldsymbol{\mu}^{(2)}, \boldsymbol{\sigma}^{(2)})$	0.0080	0.0090	0.0435	0.0335	0.0725	0.0720
	$(\boldsymbol{\mu}^{(3)}, \boldsymbol{\sigma}^{(3)})$	0.0075	0.0070	0.0365	0.0500	0.0780	0.0755

Table 2.2: Simulation result 2

$\mathbf{n} = (n_1, n_2, n_3, n_4, n_5, n_6)$ is a vector of unequal group sizes with $\mathbf{n}^{(1)} = (10, 12, 12, 16, 16, 20), \mathbf{n}^{(2)} = (10, 10, 10, 10, 10, 10), \mathbf{n}^{(3)} = (20, 16, 16, 12, 12, 10)$; $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ is a vector of unequal means, with $\boldsymbol{\mu}^{(1)} = (1, 1, 1, 1, 1, 1), \boldsymbol{\mu}^{(2)} = (1, 1, 1, 1, 1, 0.8), \boldsymbol{\mu}^{(3)} = (1, 1, 1, 1, 1, 0.9)$; $\boldsymbol{\sigma} = (\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2, \sigma_6^2)$ is a vector of unequal variances, with $\boldsymbol{\sigma}^{(1)} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1), \boldsymbol{\sigma}^{(2)} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.5), \boldsymbol{\sigma}^{(3)} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.3)$. "Overall" means overall test from Li, 2009 for equality of group means; "MCP" is the FGPQ-based multiple comparison procedure (proposed method);; Numbers in Table are simulated p-values.

		$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
$\mathbf{n}$	$(\boldsymbol{\mu}, \boldsymbol{\sigma})$	Overall	MCP	Overall	MCP	Overall	MCP
$\mathbf{n}^{(1)}$	$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\sigma}^{(1)})$	0.0060	0.0125	0.0445	0.0530	0.0845	0.0995
	$(\boldsymbol{\mu}^{(2)}, \boldsymbol{\sigma}^{(2)})$	0.0065	0.0140	0.0475	0.0590	0.0830	0.0925
	$(\boldsymbol{\mu}^{(3)}, \boldsymbol{\sigma}^{(3)})$	0.0070	0.0140	0.0440	0.0500	0.0855	0.0985
$\mathbf{n}^{(2)}$	$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\sigma}^{(1)})$	0.0065	0.0150	0.0470	0.0600	0.0795	0.0920
	$(\boldsymbol{\mu}^{(2)}, \boldsymbol{\sigma}^{(2)})$	0.0070	0.0140	0.0440	0.0570	0.0800	0.0875
	$(\boldsymbol{\mu}^{(3)}, \boldsymbol{\sigma}^{(3)})$	0.0060	0.0115	0.0465	0.0580	0.0835	0.0895
$\mathbf{n}^{(3)}$	$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\sigma}^{(1)})$	0.0070	0.0140	0.0480	0.0570	0.0870	0.0920
	$(\boldsymbol{\mu}^{(2)}, \boldsymbol{\sigma}^{(2)})$	0.0080	0.0185	0.0430	0.0600	0.0890	0.1050
	$(\boldsymbol{\mu}^{(3)}, \boldsymbol{\sigma}^{(3)})$	0.0080	0.0125	0.0495	0.0525	0.0965	0.0935

Chapter 3

Analysis

In this chapter, we consider a data set which is based on the Carbon Reduction Commitment Energy Efficiency Scheme (CRC) in the UK (for more details see Department of Energy and Climate Change (2014)). This government scheme requires all large organizations which meet the specified participation criteria to report their total energy use every year and pay a levy for each tCO_2 (ton of carbon dioxide or equivalent gasses) emitted. Also reported by many organizations is their energy intensity (Emissions over Revenue) each year, commonly referred to as the "Growth Emissions". In the first year, they also report their industry classification, the percentage of emissions covered by Automated Meter Reading (AMR), and the percentage of emissions covered by an Energy Management Standard (EMS), as well as other variables which we will not use in this analysis. We will use the first two years of the scheme, Year 1 (Financial Year 2010 / 2011) and Year 2 (Financial Year 2011 / 2012) as those data set are currently available.

Organizations with better AMR and EMS coverage were given higher positioning in the public Performance League Table for the first two years of the scheme (when no historical data was available), as it was felt that these two measures indicated that an organization was actively working to improve their energy efficiency. We are therefore interested in finding whether organizations which had higher percentages of AMR and EMS coverage achieved a better savings ratio in Year 2.

Our response variable is the ratio of emissions in Year 2 versus Year 1, $X =$ Emissions Year 2/Emissions Year 1. A value of 1 therefore indicates no change, a value > 1 indicates higher emissions in Year 2, and values < 1 indicate a saving in Year 2.

In this section we will first outline the CRC scheme in more detail and discuss the variables of interest, as well as show why this dataset can be justified to be modelled using a log-normal distribution. We will then apply the methods outlined in Section 2 to carry out the following analysis on this data set:

Equivalence of Absolute and Growth Emissions We test whether the two response variables Absolute Emissions and Growth Emissions are equivalent using the test for equivalence of bivariate log-normal means

Overall Saving Test whether the savings in Year 2 compared to Year 1 are statistically significant

AMR We test whether organizations with some AMR have a better savings ratio than those with no AMR, and also further subdivide organizations into four groups depending on their percentage of AMR and carry out pairwise comparison among the three groups.

EMS Similar to the AMR analysis, we test the mean savings ratios of organizations with no EMS vs some EMS, and the equivalence among four groups with different EMS.

3.1 CRC Energy Efficiency Scheme

The CRC Energy Efficiency Scheme (often referred to as simply “the CRC”) is a mandatory scheme aimed at improving energy efficiency and cutting emissions in large public and private sector organisations in the United Kingdom (Department of Energy and Climate Change, 2014). These organizations are responsible for around 10 percent of the UKs greenhouse gas emissions - there are currently just over 2000 organizations with reporting obligations.

The CRC affects large public and private sector organizations across the UK. Participants include supermarkets, water companies, banks, local authorities and all central government departments. Qualifying organizations are required to report their annual energy consumption to the Environment Agency (or the Scottish Environmental Protection Agency or Northern Irish Environment Agency depending on jurisdiction), together with a number of other metrics.

The energy consumption in MWh is then used to calculate tCO_2 based on the UK’s energy mix for each year, and so calculate the CO_2 emissions indirectly generated by each organizations energy use. The organization is then required to pay a carbon levy per tCO_2 emitted, thus providing a direct financial incentive to reduce energy consumption.

In addition to this levy, the Environment Agency also publishes an annual Performance League Table (PLT), which ranks organizations according to three metrics:

Absolute Metric

The current year’s energy consumption compared to the historical average from previous years

Growth Metric

The current year’s energy consumption per £1m revenue, compared to the historical average of the same

Early Action Metric

The percentage of energy consumption covered by AMR, and the percentage of energy consumption covered by the Carbon Trust Standard (or equivalent) Energy Management Scheme (EMS)

The weights of these three metrics for the overall rank changes as the scheme progresses. The Early Action Metric is only used for the first two years when historical data is not available. It was introduced to incentizive the uptake of automated systems of meter readings which would enable organizations to more closely monitor their energy use with the aid of a real-time computer system, and the adoption of an energy management standard which would allow them to identify and act upon energy saving opportunities throughout the organization.

The data used in this thesis is based on the publically available Performance League Table (PLT). The PLT for each year can be downloaded from the Environment Agency's website as an XML file (Environment Agency, 2011, 2012). Please see Appendix A for full details on how these XML files were used to create a CSV file with the data set to be analyzed. We are interested in the following variables:

Absolute Emissions

The absolute emissions in tCO_2, calculated by the Environment Agency by converting the reported energy consumption in MWh to tCO_2 using each year's Energy Mix for the United Kingdom. The savings ratio are the current year's emissions divided by the rolling historical average emissions.

Growth Emissions

What is usually referred to the Growth Emissions are the energy intensity, given by the year's Absolute Emissions divided by the revenue, given in units of £1m. The resulting value is tCO_2 per £1m revenue, and again the saving ratio is the current year's growth emissions divided by the historical average.

SIC Code

The Standard Industry Code (SIC) for the organization, in the following format: N.23.12, where N denotes the overall industry, the first number group (23) denoted the industry subsector, and the final number group (12) denotes the specialization. This is only available for private sector organizations.

AMR

The percentage of emissions in that year covered by AMR. This is a required field for all organizations, and takes values between 0 - 100. Around 34% of the organizations had no AMR in Year 1. This variable is only reported in the first two years as it is part of the Early Action Metric which has no impact on the PLT from Year 3 onwards.

EMS

The percentage of emissions in that year covered by an EMS. As with AMR, this is a required field and takes values between 0 - 100. Around 65% of the organizations had no EMS in Year 1. Similar to AMR, it is only reported in the first two years.

There were 2278 organizations which reported in Year 1. However, we only use the data for organizations which reported both in Year 1 and Year 2 (if an organization did not meet the qualification criteria in one of the years they were not required to report), and which reported both their Absolute and their optional Growth Emissions in both years. This brings the total number of useable observations to 1314. The reason we restrict this is as part of our analysis will be on the equivalence of the Absolute and Growth Emissions variables.

For future publication years, the scheme has been simplified and there will no longer be a full PLT published. Instead, a spreadsheet will be made available each year which will contain only the Absolute Emissions, Historical Average of Absolute Emissions, and savings percentage. See the Future Work section for details of the implication of this on our analysis.

3.2 Data Source and Structure

The Performance League Table for each year is available to download as an XML file from the Environment Agency's website. For detailed instructions on how these two XML files were converted into one .csv file which is used in this analysis, see Appendix A.

Most of these 32 variables are not of interest to this analysis, and we keep the following variables in the final data file:

1. **Registrant Number:** The unique identifier for each organization
2. **Registrant Name:** Company / Organization name
3. **Trading Name:** If different from the Registrant Name
4. **SIC Code:** The Standard Industrial Classification Code classifies private companies into industries
5. **Company Number:** The official Company Number, only available for private sector organizations
6. **Public Body Type:** Free-text field giving details of the type of public sector organization

7. **Industry:** This variable is not an original one in the XML file, but is simply the letter from the SIC code, that is the highest level industry categorization. The Public sector organizations which do not have an SIC code were assigned the value "AA"

8. **Absolute Emissions 1:** The Absolute Emissions for Year 1, in tCO_2

9. **% Emissions Covered by AMR 1:** The percentage of total emissions covered by voluntary AMR in Year 1

10. **% Emissions Covered by EMS 1:** The percentage of total emissions covered by the Carbon Trust Standard or an equivalent EMS, in Year 1

11. **Emissions per million £Turnover 1:** The Growth Emissions for Year 1, in tCO_2 / £1m

12. **Absolute Emissions 2:** The Absolute Emissions for Year 2, in tCO_2

13. **% Emissions Covered by AMR 2:** The percentage of total emissions covered by voluntary AMR in Year 2

14. **% Emissions Covered by EMS 2:** The percentage of total emissions covered by the Carbon Trust Standard or an equivalent scheme (EMS), in Year 2

15. **Absolute Achivement Result:** The percentage saving in Absolute Emissions of Year 2 compared to Year 1. A positive value indicate a saving, a negative value indicate higher emissions in Year 2

16. **Emissions per million £Turnover:** The Growth Emissions for Year 1, in tCO_2 / £1m

17. **Growth Achivement Result:** The percentage saving in Growth Emissions of Year 2 compared to Year 1. A positive value indicate a saving, a negative value indicate higher emissions in Year 2

Our response variables of interest are the ratio of Absolute Emissions and the ratio of Growth Emissions, that is:

$$\begin{aligned} X_{abs} &= \text{Absolute Emissions}_2/\text{Absolute Emissions}_1 \\ X_{growth} &= \text{Growth Emissions}_2/\text{Growth Emissions}_1 \; . \end{aligned} \tag{3.1}$$

We will also use the percentage values of AMR and EMS from Year 1 as explanatory variables.

The variables included in the original XML files which were not used in this analysis include the details of scores for placement in the Performance League Table as the PLT positioning is not of interest to us, as well as details of emissions covered by renewable energy generation, as very few organizations reported these details. There were also fields for participant comments, and yes/no tickboxes for questions about the organizational structure which were not of interest to our analysis.

3.3 Log-Normal Assumption and Grouping

The data set downloadable from the Environment Agency's website contains a variable for the percentage savings between Year 2 and the historical average (Year 1), for both Absolute and Growth Emissions. However, testing for normality on both of these variables separately using the Shapiro-Wilks test gives us a p-value < 0.0001 for both variables, indicating we should reject the null hypothesis that they are normally distributed.

As some of the these percentages are negative, we instead use the ratio of emissions between Year 2 and Year 1, to allow us to carry out a log-transformation. We construct two new response variables as per Equation 3.1, the ratio of Absolute Emissions X_{abs} and the ratio of Growth Emissions X_{growth}. However, even after log-transformation these variables are not normally distributed, as the Shapiro-Wilks test

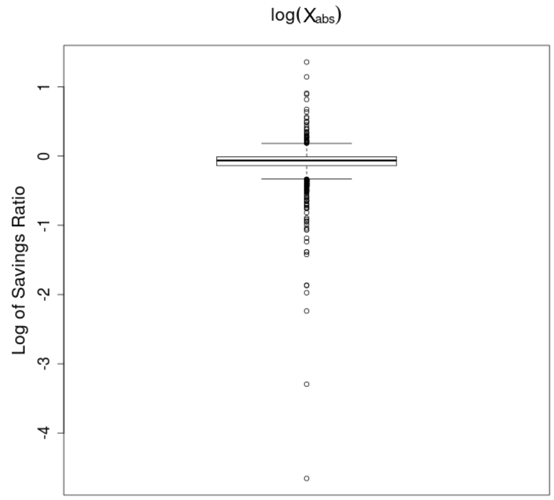

Figure 3.1: Boxplot of $log(X_{abs})$

This ratio show a much higher variance / spread than expected from a normal distribution

gives p-values less than 0.0001 for both $Y_{abs} = log(X_{abs})$ and $Y_{growth} = log(X_{growth})$. Looking at the boxplot in Figure 3.1, we see that the data does not appear to be normally distributed.

To reduce variability, we group the organizations together in Industry groups, by using the first letter in the Standard Industry Classification (SIC) code. The letter denotes the main industry grouping, and using the UK SIC scheme the private organizations in our dataset is divided into 13 groups, with one additional group for all Public sector organizations (as they do not have an SIC code).

However, out of these 13 groups there are five groups with fewer than 10 organizations. These groups therefore exhibit large variations, and in order to ensure normality we group the organizations from these five groups into one industry classification, “Other”. This ensures that all industry groups have >10 observations each. We therefore have 10 groups in total whose means are representative of their ungrouped observations (8 from SIC letters, one for Public organizations, and one for "Other" industries) - see Figure 3.3.

Carrying out the Shapiro-Wilks test for normality on the 10 data points after industry grouping has been carried out gives the following p-values:

$$Y_{abs} = log(X_{abs}) : W = 0.8834, \text{p-value} = 0.1428$$
$$Y_{growth} = log(X_{growth}) : W = 0.9628, \text{p-value} = 0.8174\ .$$

Hence after grouping by industry, our basic assumption of log-normality holds as both p-values > 0.05. Although the data also passes the Shapiro-Wilks normality test for the non-log-transformed variable after grouping (and hence could be analyzed using standard statistical methods), it is more appropriate to use the log-normal distribution as the data fullfills the three properties outlined in the Methodology section:

1. **Positive values**: As energy usage can take only positive values the ratio of two year's energy usage is always positive.

2. **Volatility**: As we have explained above, the variance of the data is larger than what is expected from a normal distribution.

3. **Annual Savings Ratio**: The ratio of energy savings for year t, defined as $X_t = \text{Emissions}_t/\text{Emissions}_{t-1}$, is similar to the rate of return $R_r = \frac{S_t}{S_{t-1}}$ in the stock price example. Hence following the same argument as in Section 2.1, if we were to analyse the energy savings over several years a log-normal model could be appropriate.

At the time of writing this thesis, only two years of energy usage data is available and hence the third point, about modelling several years of savings, seems moot - however as the CRC scheme is intended to continue for some time, if the data is

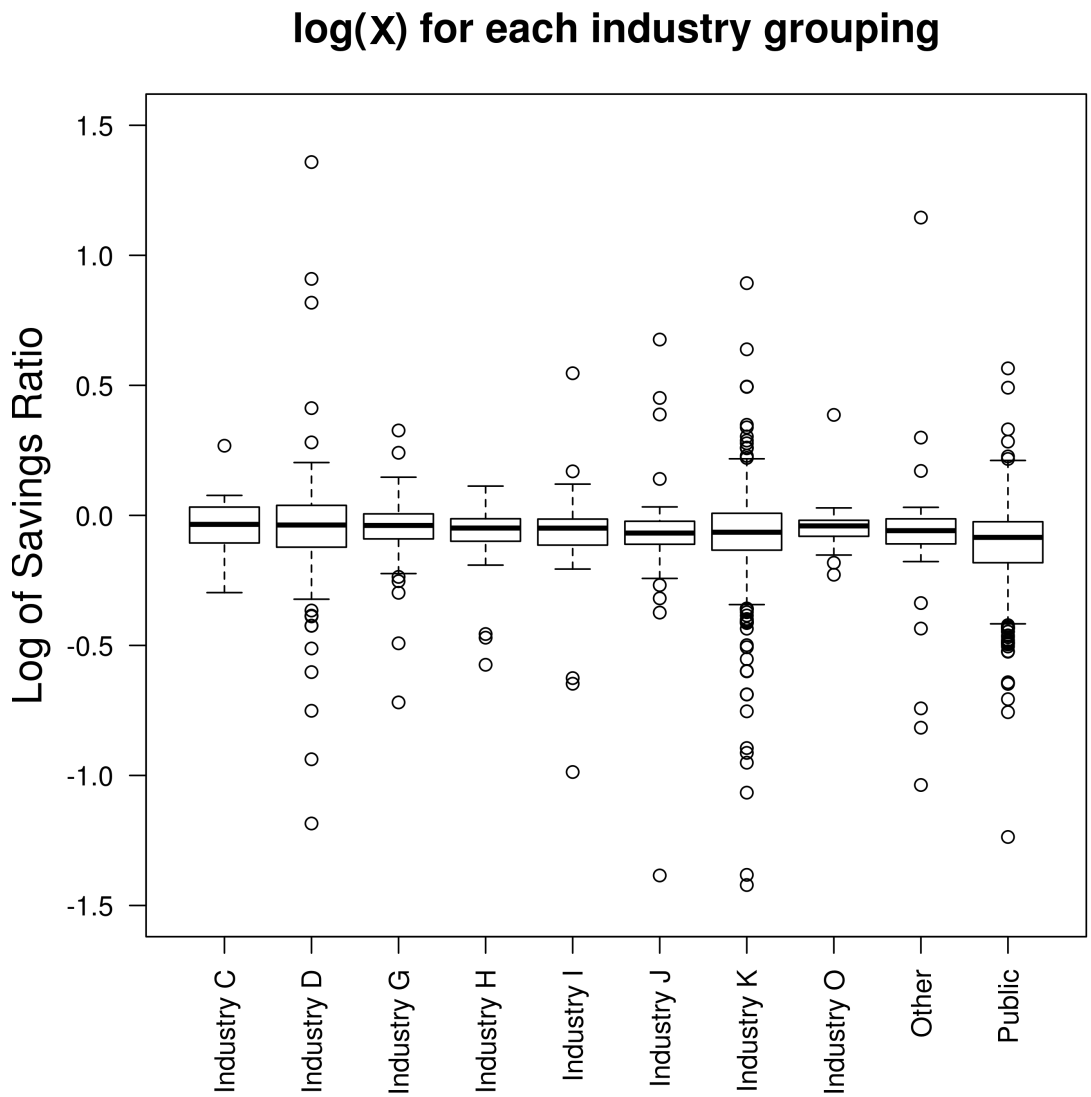

Figure 3.3: Boxplot of $log(X_{abs})$ for each industry grouping

We see that all industry groups have similar means and variances

analyzed using log-normal methods now it can easily be generalized to incorporate additional data when it becomes available, whereas a normal model might be more restricted to a one-off analysis.

In all sections below the variables Y and X will indicate the variables grouped by industry.

3.4 Equivalence of Absolute and Growth Emissions

Before investigating the savings ratios for different groups of AMR and EMS, we need to ensure we use the correct response variables. There are two possible variables, the ratio of Absolute Emissions, and the ratio of Growth Emissions (energy intensity). We wish to test whether these two response variables are equivalent, and so model them as a bivariate log-normal distribution, using the notation from Section 2.4 :Y_1 = Y_{abs} and Y_2 = Y_{growth}, to test the null hypothesis:

$$H_0 : \eta_{abs} = \eta_{growth} \quad \text{vs} \quad H_\alpha : \eta_{abs} \neq \eta_{growth} \ , \tag{3.2}$$

where η_{abs} is the log-normal mean of $Y_{abs} = log(X_{abs})$, and η_{growth} is the mean of the log-normal variable $Y_{growth} = log(X_{growth})$. Using the method outlined in section 2.4 for testing the equivalence of bivariate log-normal means, we wish to construct a confidence interval for θ, where $\theta = \eta_{abs} - \eta_{growth} = (\mu_{abs} - \mu_{growth}) + \frac{1}{2}(\sigma_{abs} - \sigma_{growth})$.

We know from Section 3.5 that Y_{abs} and Y_{growth} are log-normally distributed. In order to construct the test statistic T from Eq (2.10), we use the function *bivariate-comparison(x, z, m, α)*, where we set $x = Y_{abs}$, $z = Y_{growth}$, the number of Monte Carlo iterations $m = 5000$ and the significance level $\alpha = 0.05$.

The output is the GCI for θ, which is (-1.834, 0.444). Therefore, the confidence interval for the ratio of the bivariate means in the units of the ratios is ($e^{-1.835}$, $e^{0.424}$) = (0.159, 1.559). As this confidence interval comfortably includes 1, we conclude we cannot reject the null hypothesis in Eq (3.2) that the ratio of the means of Absolute and Growth Emissions are not significantly different.

We will therefore proceed with the analysis of comparisons of means below using only the Absolute Emissions variable. There are three reasons for this:

1. As we have just shown, the means for the two response variables Absolute and Growth Emissions are not significantly different.
2. The variance of the Growth variable is very high, so the lower variance of the Absolute variable allow us to better find the main effects of the explanatory variables.
3. As of Year 3, the Growth variable will no longer be reported, hence in order to make it easy to incorporate future data into this analysis only Absolute Emissions should be used.

3.5 Overall Significance of Savings Ratios

In this section we will investigate whether the savings in Year 2 are statistically significant. As outlined above, as the response variable we will use the ratio of Absolute Emissions, Y_{abs}. We wish to determine whether energy savings ratio Y_{abs_2}/Y_{abs_1} is lower than the no change scenario, at a significance of $\alpha = 0.05$.

As we know Y_{abs} can be modelled as being log-distributed, we can use the methodology in section 2.4 to test out the following hypothesis, where η_{abs} is the log-normal mean of Y_{abs}. We set the constant $\eta_0 = 0$, as that would give a savings ratio of $e^{\eta_0} = 1$, i.e. no change between Year 1 and Year 2:

$$H_0 : \eta_{abs} \geq 0 \quad \text{vs} \quad H_\alpha : \eta_{abs} < 0 \ . \tag{3.3}$$

Using Eq (2.8) to construct the generalized test statistic T, we can construct a confidence interval for η_{abs}, and a p-value for the null hypothesis in Eq (3.3). This is done by using the function $CIFunction(a, \eta_0, m, \alpha)$ with $a = Y_{abs}$, $\eta_0 = 0$, $m = 5000$ and $\alpha = 0.05$.

The output of this function is a 95% upper confidence interval for η_{abs} of -0.0413 and a p-value for the null hypothesis in Eq (3.3) of $p = 0$ (which since we used $m = 5000$ equates to $p < 0.002$). Hence as the p-value is far below the significance level of $\alpha = 0.05$, and the upper confidence level is below 0, we conclude that the overall Absolute Emissions for all participants were significantly lower in Year 2 than in Year 1.

As the overall savings are significant, we can continue our analysis of the difference of organizations with AMR and EMS.

3.6 Automated Meter Reader

In the first year of the scheme, participants were asked to submit information on the percentage of their total emissions which was covered by AMR. AMR refers to technologies which automatically reports energy usage data to a central location without the need for manual meter readings, and which allows an organization to identify areas where energy usage could be streamlined, and to get a real-time picture of the current energy usage. This variable was included to allow a Performance League Table (PLT) to be published for Year 1, as it was felt that uptake of AMR would allow organizations to more efficiently monitor their energy usage, and hence identify saving options.

This variable takes a value between 0 - 100%, with approximately one third of the reporting organizations having no AMR. In order to investigate the effects of AMR on energy savings in the first year of the CRC, we will carry out a comparison of the mean Absolute Emissions savings ratio of organizations with different levels of AMR, allocated into groups.

We first carry out a comparison of two groups, organizations with no AMR (group 1: G_1), and organizations with AMR>0 (group 2: G_2), to see if organizations with some AMR had a better savings percentage (lower ratio) than those with no AMR. This is therefore a comparison of two independent log-normal means, and we will proceed to find the generalized upper confidence limit for Eq (3.4), using the methods outlined in Section 2.8.

$$H_0 : \eta_{G_2} \geq \eta_{G_1} \quad \text{vs} \quad H_\alpha : \eta_{G_2} < \eta_{G_1} \ . \tag{3.4}$$

To construct the two groups G_1 and G_2, we subdivide each Industry group in those organizations with no AMR, and those with some AMR. For example, for the organizations where Industry = C, we divide them into those with no AMR G_{1_C} and those with some AMR G_{2_C}, and the mean absolute savings ratio of these organizations respectively is the mean of G_{1_C} and G_{2_C}. Therefore η_{G_1} is the log-normal mean of G_{1_j}, where j are the 10 industry groupings.

To carry out the comparison of independent log-normal means as outlined in Section 2.8, to test the hypothesis in Eq (3.4), we use the custom function *comparsion*(η_a, η_b, m, α), where η_a is the log-normal mean of G_2 (with some AMR), η_b is the log-normal mean of G_1 (no AMR), number of iterations $m = 5000$ and a significance level of $\alpha = 0.05$.

The output of this function gives an upper level confidence interval value of 0.085, and a p-value of 0.8968. As the upper confidence interval is above 0, and $p > \alpha$, we conclude that we cannot reject the null hypothesis H_0 in Eq (3.4), and hence we see that organizations with some AMR did not achieve a better savings ratio compared to those with no AMR.

Next, we wish to see if there is a difference in saving ratios for different percentages of AMR. We therefore divide each industry group into four segments as follows:

- G_1: AMR % = 0
- G_2: $0 <$ AMR % $\leq 33\%$
- G_3: $33 <$ AMR % $\leq 66\%$
- G_4: $66 <$ AMR % $\leq 100\%$

For G_2, there were no industries from group C which satisfied this criteria, and for G_4 only one company from industry group H, so this observation was removed as it provided too large variance for the log-normality test. With these two adjustments, all four groups pass the Shapiro-Wilk test for log-normality at $\alpha = 0.05$. Thus we have four groups with different log-normal mean saving ratios (η_i, $i = 1, 2, 3, 4$), and we wish to carry out all pairwise comparisons and construct confidence intervals for the differences $\eta_1 - \eta_2, \eta_1 - \eta_3$ etc. The GCI we construct will test the null hypothesis:

$$H_0 : \eta_i - \eta_j = 0 \quad \text{vs} \quad H_\alpha : \eta_i - \eta_j \neq 0 \tag{3.5}$$

where $i, j = 1, 2, 3, 4$ and $i \neq j$.

Using the methods from section 2.6, we use the function *severalmeans(data, m, α)* where the array *data* are the groups $G_{1,2,3,4}$, number of iterations $m = 5000$ and the significance level $\alpha = 0.05$. The output from the function is as follows:

From this table we can conclude that the only difference that was statistically significant at $\alpha = 0.05$ was that between G_1 and G_4, that is organizations with an AMR percentage $> 66\%$ had a higher ratio than those with no AMR. All other

i	j	Lower CI	Upper CI
1	2	-0.963	0.047
1	3	-0.846	0.052
1	4	-0.131	-0.006
2	3	-0.058	0.074
2	4	-0.106	0.017
3	4	-0.111	0.006

comparisons included 0 in their confidence interval hence we cannot reject the null hypothesis that there is no difference among them.

As mentioned previously, AMR (and EMS) were included in the CRC reporting as it was felt that these measures would allow organizations to easier identify energy saving opportunities. This analysis shows no difference in energy savings between groups with AMR and groups with no AMR, and in fact indicates that organizations with high AMR performed worse than those with no AMR. The data that is currently available is only for one ratio (from the first two years of the scheme), and hence there are two additional factors to take into consideration when interpreting these results: first, organizations with high AMR are more likely to have worked on improving their energy efficiency prior to the first CRC reporting year, and secondly energy efficiency improvements are subject to diminishing returns, as discussed by for example Jaffe and Stavins (1995) and Arimura et al. (2011).

As can be seen from a case-study of AMR installation by Leicester City Council (Ferreira et al., 2007), AMR allows organizations to identify and act on easy saving opportunities, the so-called "low-hanging fruit". As achieving high AMR coverage usually takes some time (in the case of Leicester City Council they had at the time of the case-study been installing AMR for 5 years and did not yet have full coverage), it is likely that organizations with high AMR started installing new meters several years prior to 2010, and hence have already acted on these easier saving opportunities. Organizations with no AMR however, are less likely to have seriously tried to improve their energy efficiency before having an additional financial incentive to do so in the form of the CRC scheme, with the result that organizations with high AMR have a worse savings ratio between 2010 and 2011, whereas organizations with no AMR could act on the easier saving opportunities before the diminishing returns took effect.

It is therefore of interest to carry out this analysis again with a few more years worth of data, when all organizations have more evenly matched starting points and the early adopters of energy efficiency aren't "penalized" as described above. Due to the multiplicative property of the log-normal distribution, this means we could carry out the same analysis as above with the ratio from future years multiplied.

3.7 Energy Management Systems

An EMS, in the context of energy efficiency, is an organizational-level system which sets out to help an organization achieve energy efficiency, using specific procedures and methods. It also includes systems for continual improvement and monitoring, which will spread awareness of energy efficiency throughout an entire organization. Under the CRC, an organization was considered to be using an EMS when it was certified under an Energy Management Standard, such as the Carbon Trust Standard (Carbon Trust, 2014) or an equivalent scheme.

The analysis here will take the same format as for AMR - we first compare no EMS to some EMS, then we divide the organizations into four groups depending on their % level of EMS and do a pairwise comparison.

First, we divide each industry group into those organizations with no EMS and those with some EMS:

- F_1: EMS % = 0
- F_2: EMS % > 0

Carrying out the Shapiro-Wilk test on $log(F_1)$ gives a p-value of 0.996, and for $log(F_2)$ it is 0.049 - thought technically it fails the log-normality test at $\alpha = 0.05$ it is close enough that we will go ahead and analyze these groups using a log-normal model. We wish to test the hypothesis:

$$H_0 : \eta_{F_2} \geq \eta_{F_1} \quad \text{vs} \quad H_\alpha : \eta_{F_2} < \eta_{F_1} \ . \tag{3.6}$$

To compare the two groups using the methods in Section 2.5, we as before make use of the custom function *comparion(*η_a*,* η_b*,* m*,* α*)*, where η_a is the log-normal mean of F_2, η_b is the log-normal mean of F_1, number of iterations $m = 5000$ and a significance level of $\alpha = 0.05$.

The upper level confidence interval for Eq (3.6) is given as 0.0935, with a p-value of 0.944. Hence we cannot reject the null hypothesis, and conclude that, as with AMR, there is no difference in savings ratios between organizations with no EMS and those with some EMS.

We then proceed to test the differences among the four groups:

- F_1: EMS % = 0
- F_2: 0 < EMS % $\leq$ 33%
- F_3: 33 < EMS % $\leq$ 66%
- F_4: 66 < EMS % $\leq$ 100%

All groups but F_3 passes the Shapiro-Wilk test of log-normality, but as the additional variance in F_3 is due to one single outlier (from Industry = P), we will continue and analyze all pairwise comparisons using the log-normal model. We then wish to test the null hypothesis:

$$H_0 : \eta_i = \eta_j \quad \text{vs} \quad H_\alpha : \eta_i \neq \eta_j \ , \tag{3.7}$$

where $i, j = 1, 2, 3, 4$, and also $i \neq j$. Using the function *severalmeans(data, m, α)* where the data array contains the groups $F_{1,2,3,4}$, the number of iterations $m = 5000$ and a significance level $\alpha = 0.05$, we see the following output for all pairwise comparisons:

i	j	Lower CI	Upper CI
1	2	-0.051	0.112
1	3	-0.127	0.089
1	4	-0.098	0.011
2	3	-0.177	0.077
2	4	-0.163	0.014
3	4	-0.137	0.088

As none of these confidence intervals exclude 0, we cannot reject the null hypothesis that the log-normal mean for all four groups are the same. Hence, there is no evidence that different % levels of EMS results in different savings ratios. As with the AMR analysis, the assumption we set out to test (that higher percentages of EMS coverage would result in higher savings ratios) is subject to the same restrictions as for AMR, where organizations with higher percentages have already acted on the easier savings opportunities and hence see diminishing returns.

Chapter 4

Conclusions & Discussion

Log-normal distribution is widely used to describe the distribution of positive random variables that exhibit skewness in biological, medical, economical and social studies. For example, the CRC energy efficiency scheme is a mandatory energy usage reporting scheme for large organizations in the UK. After grouping by industry, data was shown to be log-normally distributed. Simultaneous confidence intervals for certain log-normal parameters are useful in pharmaceutical and other statistics. Our research addresses the simultaneous confidence interval problem for data from several log-normal distributions under heteroscedasticity and unequal group sizes.

In this thesis, we first gave an overview of the GCI approach and reviewed three tests: GCI for inference of one mean, GCI for testing equivalence of bivariate response variables, and GCI for the difference between two independent log-normal means. Based on the concepts of FGPQ (a subclass of GCI), we then proposed simultaneous confidence intervals for ratios of the means from k log-normally distributed data under heteroscedasticity and unequal group sizes. We have proved that the constructed confidence intervals have correct asymptotic coverage. A computing algorithm was proposed to construct the confidence intervals. Simulations showed that the simulated p-values are close to the nominal level. The proposed methods work well even for small (~ 10) samples. This research together with overall tests by Li (2009) provide a solution of inference on several log-normal distributions.

The existing GCI methods and the proposed methods were applied to analyze data from the CRC scheme for financial Year 2010 / 2011 (Year 1) and 2011/2012 (Year 2). Using the bivariate log-normal GCI test, the absolute emissions was shown to be a representative response variable. We also concluded that the overall energy saving for Year 2 was significant compared to Year 1. After dividing organizations into groups by their percentage of AMR and EMS coverage, and using the proposed simultaneous pairwise comparison, we have found that there is no significant difference among the groups, with the exception that organizations with high AMR ($> 66\%$) performed worse than organizations with no AMR. Organizations that had already been working on their energy efficiency are more likely to already have AMR installed and therefore ended up in the higher groups of AMR %, whereas organizations which did not carry out energy efficiency work prior to the start of the scheme are unlikely to have installed any AMR. Therefore the no AMR group showed a better savings ratio for Year 2 compared to the high AMR group, as they were able to put in effect the easier energy saving opportunities, which the high AMR group organizations had already carried out prior to 2010.

The CRC scheme will continue collecting, and publishing energy usage data in future years in a slightly different format. For future publication years, the scheme has been simplified, and a spreadsheet will be made available each year containing only the Absolute Emissions, Historical Average of Absolute Emissions, and savings percentage. We could integrate the new data into the existing data source file, and extend our analysis to the saving ratios for future years using the existing R code. We also intend to address the longitudinal analysis problem using continuous-time Markov models or time series models when more future years' CRC data is available.

Our research on simultaneous pairwise comparisons using GCI is done in the context of log-normal model, but it can also be extended to other distributions where confidence intervals based on sufficient statistics are not available.

Appendices

Appendix A

Data Source

The data is available as two XML files, one for each year of the scheme, to download from the Environment Agency's PLT website: (Environment Agency, 2011), (Environment Agency, 2012).

The XML files were imported in Excel using the XML importing tool, and the following steps were carried out:

1. Due to the three free text boxes for Participant Comments, each organization was listed three times when the XML import was completed - two of these were deleted for each organization.

2. The data from the two separate files were copied into a new file. Using the VLOOKUP function in Excel, the unique CRC identifier for each organizations was used to ensure the data from the same CRC Identifier were copied to the same row.

3. The column names for the Year 2 data had a "2" added at the end of the variable name - so "crcEmissions" and "crcEmissions2" refer to the absolute emissions from Year 1 and Year 2 respectively.

4. Any organizations which did not have a value in both "emissionsPerMillionPoundsTurnover" and "emissionsPerMillionPoundsTurnover2" were deleted, as that indicated an organization which did not report their Growth Emissions in both years (Absolute Emissions were mandatory to report).

5. Superfluous columns not used in the analysis were removed.

6. A new column "Industry" was added, which contained the first letter from the SIC code if that was available, and the letter "P" (for Public) if there was no SIC code. In addition, due to the low number of organizations in industry groups A, E, F, L, N and M, these were all given Industry code A, indicating "other" industry.

7. A new column "Private" was created, where an organizations was given the value 1 if it was a private organization, and 0 if it was not. This data did not end up being used in the final analysis.

8. The file was saved in the .csv format, allowing it to be easily imported into R

The final columns / variables in the .csv file are, along with their numbering (used for referencing the column number in some of the below code):

1. registrantNumber

2. registrantName

3. tradingName

4. sicCode

5. companyNumber

6. publicBodyType

7. Industry
8. Private
9. crcEmissions
10. percentageOfEmissionsCoveredByVoluntaryAMR
11. percentageOfEmissionsCoveredByCarbonTrustStandard
12. emissionsPerMillionPoundsTurnover
13. crcEmissions2
14. percentageOfEmissionsCoveredByVoluntaryAMR2
15. percentageOfEmissionsCoveredByCarbonTrustStandard2
16. absoluteAchievementResult2
17. emissionsPerMillionPoundsTurnover2
18. growthAchievementResult2

Appendix B

R Codes

B.1 GCI functions overview

There were four custom functions defined in R which were used to analyze the data in Section 3. What follows is a plain text explanation of the R functions created for calculating the GCI's, based on the Algorithms outlined in Section 2. The full code is available below.

$CIFunction(a, \eta_0, m, \alpha)$: The input variables are:

a The variable to be tested

$\boldsymbol{\eta_0}$ A constant, for testing $\eta_a < \eta_0$

m The number of Monte Carlo simulations

$\boldsymbol{\alpha}$ Significance level for Confidence Intervals and p-values

Following the structure of Algorithm 1, first various statistics are calculated for $log(a)$, and these are then used in Eq (2.8) to create the test statistic T, using m

Monte Carlo iterations for the randomly generated values of Z and U^2. The output is the upper $100(1-\alpha)\%$ confidence interval to the for η_a.

$bivariatecomparison(x, z, m, \alpha)$: The input variables are:

x & z The bivariate variables to be tested for equivalence

m The number of Monte Carlo simulations

$\boldsymbol{\alpha}$ Significance level for constructing the Confidence Intervals

Using Algorithm 2, summary statistics of $log(x)$ and $log(z)$ are calculated, and these are then used in Eq (2.10) to create the test statistic T, using m Monte Carlo iterations for the randomly generated values of Z and U^2. The output is the confidence interval to the $100(1-\alpha)$ percent for T.

$comparison(a, b, m, \alpha)$: The input variables are:

a The variable to be tested for $a < b$

b The variable to be tested for $a < b$

m The number of Monte Carlo simulations

$\boldsymbol{\alpha}$ Significance level for constructing the Confidence Intervals

Various statistics are first calculated for $log(a)$ and $log(b)$, which are used in a modified Algorithm 1 using Eq (2.13) and Eq (2.14) to create the test statistic T_a and T_b, using m Monte Carlo iterations for the randomly generated values of Z and U^2. The output is the two-sided confidence interval to the $100(1-\alpha)$ percent for $T = T_a - T_b$.

severalmeans$(array, m, \alpha)$: The input variables are:

array An array of all k variables for which we want a pairwise comparison, where each column in the array contains one group, and there is no other data in the array.

m The number of Monte Carlo simulations

$\boldsymbol{\alpha}$ Significance level for constructing the Confidence Intervals

For each pair k and $k-1$, the log is taken of the variable and a confidence interval for $\eta_k - \eta_{k-1}$ is constructed using Algorithm 4, with the test statistic Eq (2.26). The output is all pairwise confidence intervals for the k variables.

B.2 R code - GCI Functions

The following R code is used to define the four custom function outlined in Section B.1:

```
## CIFunction(a, eta_0, m, alpha) creates a confidence interval for a, and p-value
    for a < eta_0

# "a" is the data variable
# "eta_0" is a constant for testing a < eta_0 - to test the hypothesis of no change
    this is set to 0
# m is the number of iterations, and alpha is the significance level

CIFunction <- function(a, eta_0, m, alpha)

{
  # Calculate means and variances for the log-transformed variable
  y <- log(a)
  n <- length(a)
```

```
ybar_a <- (1/n)*sum(y)
s2_a <- (1/(n-1))*sum((y - ybar_a)^2)
s_a <- sqrt(s2_a)

T <- rep(0,m)          # Empty array for the test statistic

for (i in 1:m)
  {
    # Define Z~N(0,1), U^2 ~ ChiSq(n-1), U ~ sqrt(U^2)
    Z <- rnorm(1,0,1)
    U2 <- rchisq(1, n-1)
    U <- sqrt(U2)

    # Set T_2i = ybar - (Z/(U/sqrt(n-1))) s / sqrt(n) + 1/2 s^2 / U^2 / (n-1)
    T[i] <- ybar_a - (Z/(U/sqrt(n-1)))*(s_a/sqrt(n)) + 0.5*(s2_a/(U2/(n-1)))
    # End i loop
  }

# Confidence Intervals are given by percentiles of T
UpperCI_b <- quantile(T, 1-(alpha))

# P-value for a < eta_0 is give by the proportion of T < eta_0
K <- rep(0, m)         # Empty array for storing k
for (i in 1:m)
  {

    if (T[i] >= eta_0)
      {
      K[i] <- 1
      }

    else
      {
      K[i] <- 0
      }

  }
# (1/m) sum(K_i) is a monte carlo estimate of the generalized p-value

pvalue <- (1/m)*sum(K)
```

```
  # Output pvalue and CI's as a list, to be called with [functionname]$UpperCI and [
      functionname$pvalue]
  CIoutput <- list(UpperCI=UpperCI_b, pvalue=pvalue)
  CIoutput

}


## comparison(a, b, m, alpha) gives a confidence interval for for T = eta_a - eta_b,
     and a p-value for eta_a < eta_b

# "a" is the variable to be tested as lower than "b"
# "b" is the variable to be tested against
# m is the number of iterations, and alpha the significance level

comparison <- function(a,b, m, alpha)

  {

  # Calculate means and variances for a and b
  y_a <- log(a)
  n_a <- length(a)

  y_b <- log(b)
  n_b <- length(b)

  ybar_a <- (1/n_a)*sum(y_a)
  s2_a <- (1/(n_a-1))*sum((y_a - ybar_a)^2)
  s_a <- sqrt(s2_a)

  eta_a <- ybar_a + (s2_a/2)

  ybar_b <- (1/n_b)*sum(y_b)
  s2_b <- (1/(n_b-1))*sum((y_b - ybar_b)^2)
  s_b <- sqrt(s2_b)

  eta_b <- ybar_b + (s2_b/2)

  # Empty arrays to store the test statistics
  T_a <- rep (0,m)
  T_b <- rep (0,m)

  for (i in 1:m)
```

```
  {
    # Define Z~N(0,1), U^2 ~ ChiSq(n-1), U ~ sqrt(U^2) for a
    Z_a <- rnorm(1,0,1)
    U2_a <- rchisq(1, n_a-1)
    U_a <- sqrt(U2_a)

    # Calculate T_a

    T_a[i] = ybar_a - (Z_a/(U_a/sqrt(n_a-1)))*(s_a/sqrt(n_a)) + 0.5*(s2_a/(U2_a/(n
        _a-1)))

    # Define Z~N(0,1), U^2 ~ ChiSq(n-1), U ~ sqrt(U^2) for b
    Z_b <- rnorm(1,0,1)
    U2_b <- rchisq(1, n_b-1)
    U_b <- sqrt(U2_b)

    # Calculate T_b
    T_b[i] = ybar_b - (Z_b/(U_b/sqrt(n_b-1)))*(s_b/sqrt(n_b)) + 0.5*(s2_b/(U2_b/(n
        _b-1)))

  }

# Calculate the difference T = T_a - T_b

T = T_a - T_b

# The confidence interval for T is given by the quantiles

UpperCI <- quantile(T, 1-(alpha/2))
LowerCI <- quantile(T, (alpha/2))

# P-value is given by the proportion of T equal to or above 0

K <- rep(0, m)

  for (i in 1:m)
    {
      if (T[i] >= 0)
        {
        K[i] <- 1
        }

      else
```

```
          {
          K[i] <- 0
          }
        }

  # (1/m) sum(K_i) is a monte carlo estimate of the generalized p-value

  pvalue <- (1/m)*sum(K)

  # Output pvalue and CI as a list, to be called with [functionname]$LowerCI etc
  CIoutput <- list(LowerCI=LowerCI, UpperCI=UpperCI, pvalue=pvalue)
  CIoutput

}

## BivariateComparison(x, z, m, alpha) tests the equivalence of the bivariate
    variables x and z

# "x" and "z" are the bivariate variables
# "m" is the number of monte carlo simulations, and "alpha" is the alpha-level
    confidence

BivariateComparison <- function(x, z, m, alpha){

  # Calculate summary stats to be used below

  y_x <- log(x)
  n_x <- length(x)

  y_z <- log(z)
  n_z <- length(z)

  ybar_x <- (1/n_x)*sum(y_x)
  s2_x <- (1/(n_x-1))*sum((y_x - ybar_x)^2)
  s_x <- sqrt(s2_x)

  ybar_z <- (1/n_z)*sum(y_z)
  s2_z <- (1/(n_z-1))*sum((y_z - ybar_z)^2)
  s_z <- sqrt(s2_z)

  # Defining the matrix a

  a_11 <- sum((y_x-ybar_x)^2)
```

```
  a_12 <- sum((y_x-ybar_x)*(y_z-ybar_z))
  a_22 <- sum((y_z-ybar_z)^2)
  a_11_2 <- a_11 - ((a_12)^2)/a_22

  # Empty array to store test statistics

  T <- rep(0,m)

  for (i in 1:m)
    {
      # Generate the random values for U and Z
      U_22 <- rchisq(1, n_x-1)
      U_11_2 <- rchisq(1, n_x-2)
      Z_1 <- rnorm(1, mean=0, sd=1)
      Z_2 <- rnorm(1, mean=0, sd=1)

      # Calculate R's
      R_22 <- a_22 / U_22
      R_12 <- (a_12/U_22) - (sqrt(a_11_2 * a_22) * (Z_1/sqrt(U_11_2)) * (1/U_22))
      R_11 <- (a_11_2 / U_11_2) + (((R_12)^2)/R_22)

      # Calculate the test statistic T
      T[i] <- (ybar_x - ybar_z) - ((Z_2 / sqrt(n_x)) * sqrt(R_11 - R_12 - R_12 + R_
          22)) + (0.5 * (R_11 - R_22))
    }

  # Confidence interval of theta are upper and lower quantiles of T

  UpperCI <- quantile(T, (1-(alpha/2)))
  LowerCI <- quantile(T, (alpha/2))

  # Output the two CI's in a list

  output <- list(UpperCI=UpperCI, LowerCI=LowerCI)

}


## severalmeans(array, m, alpha) gives us a confidence interval for each pairwise
    comparison of the k variables in "data"

# "data" is an array where each entry is a variable containing the data from each of
     the k groups
```

```
# m is the number of iterations, and alpha the significance level

severalmeans <- function(data, m, alpha)

  {

  k <- length(data)      # k = the number of variables to be compared

  # k*(k-1) / 2 gives the total number of combinations of comparisons
  # Empty arrays for storing eta_i and eta_j for each combnation
  J <- rep(0,(k*(k-1))/2)
  I <- rep(0,(k*(k-1))/2)

  # Empty arrays for storing all eta_i - eta_j confidence intervals
  UpperCI <- rep(0,(k*(k-1))/2)
  LowerCI <- rep(0,(k*(k-1))/2)

  # The function will compare eta_j - eta_i. j can take values from 1 to k-1
  for (j in 1:(k-1))
    {

    # For each given j, i can take values from j+1 to k. For example when j = 2 and
        k = 4, we want to compare eta_2 - eta_3 and eta_2 - eta_4.
      for (i in (j+1):k)
        {

          # Take logs and calculate n of the variables
          y_a <- log(data[[j]])
          n_a <- length(data[[j]])

          y_b <- log(data[[i]])
          n_b <- length(data[[i]])

          # Calculate means and variances
          ybar_a <- (1/n_a)*sum(y_a)
          s2_a <- (1/(n_a-1))*sum((y_a - ybar_a)^2)
          s_a <- sqrt(s2_a)

          eta_a <- ybar_a + (s2_a/2)

          ybar_b <- (1/n_b)*sum(y_b)
          s2_b <- (1/(n_b-1))*sum((y_b - ybar_b)^2)
          s_b <- sqrt(s2_b)
```

```
eta_b <- ybar_b + (s2_b/2)

#  Calculate eta_ij
eta_ij = ybar_a - ybar_b + (((n_a - 1) / (2 * (n_a - 3))) * s2_a) - (((n_b
     - 1) / (2 * (n_b - 3))) * s2_b)

# Calculate V_ij
V_ij = (((n_a - 1) / (n_a * (n_a - 3))) * s2_a) + ((((n_a - 1)^2) / (2 *
    ((n_a - 3)^2) * ((n_a - 5)^2))) * s2_a^2) + (((n_b - 1) / (n_b * (n_b
    - 3))) * s2_b) + ((((n_b - 1)^2) / (2 * ((n_b - 3)^2) * ((n_b - 5)^2))
    ) * s2_b^2)

# Prepare the dataframes for m number of simulations
R_theta <- rep(0,m)
R_xi <- rep(0,m)
Q <- rep(0,m)

  # Carry out m iterations of the test statistic
  for (z in 1:m)
    {

      # Generate values for Z~N(0,1) and U^2_k ~ ChiSq(n_k-1)
      Z_a <- rnorm(1,0,1)
      U2_a <- rchisq(1, n_a-1)
      U_a <- sqrt(U2_a)

      Z_b <- rnorm(1,0,1)
      U2_b <- rchisq(1, n_b-1)
      U_b <- sqrt(U2_b)

      # Calculate R_theta
      R_theta[z] = ybar_a - ybar_b - sqrt((n_a - 1)/n_a) * ((s_a * Z_a) /
          (U_a)) + sqrt((n_b - 1)/n_b) * ((s_b * Z_b) / (U_b)) + (((n_a -
          1) * s2_a) / (2*U2_a)) - (((n_b - 1) * s2_b) / (2*U2_b))

      # Calculate R_xi
      R_xi[z] = (((n_a - 1)*s2_a) / (n_a * U2_a))  +  (((n_a-1)^2) / (2 *
          n_a * (n_a - 3)^2))  *  (((n_a - 1) * s2_a) / (U2_a))^2  +  (((n
          _b - 1)*s2_b) / (n_b * U2_b))  +  (((n_b-1)^2) / (2 * n_b * (n_b
           - 3)^2))  *  (((n_b - 1) * s2_b) / (U2_b))^2

      # Calculate Q and q
```

```
        Q[z] = max(abs((R_theta[z] - eta_ij) / sqrt(R_xi[z])))
        q <- quantile(Q, (1-(alpha/2)))

      } # end z loop

    # The confidence interval is given by eta_ij +- (q * sqrt(V_ij)). Save
        these values in the below dataframes at the location given by the
        function [(i-j) + ((j-1)*k) - ((j*(j-1))/2)]

    UpperCI[(i-j) + ((j-1)*k) - ((j*(j-1))/2)   ] <- eta_ij + (q * sqrt(V_ij
        ))
    LowerCI[(i-j) + ((j-1)*k) - ((j*(j-1))/2) ] <- eta_ij - (q * sqrt(V_ij))
    J[(i-j) + ((j-1)*k) - ((j*(j-1))/2) ] <- j
    I[(i-j) + ((j-1)*k) - ((j*(j-1))/2) ] <- i

  }       # end i loop


 }  # end j loop

# Output J, I, UpperCI and LowerCI in the dataframe PairwiseCI and print this as
    the output of the function
PairwiseCI <- data.frame(J, I, UpperCI, LowerCI)
  PairwiseCI

}
```

B.3 R code - Industry Grouping and Misc

The data, available in .csv format following the procedure in Appendix A, was read into R, which was also used to produce the boxplots in Section 3.

Producing the p-values for testing for log-normality on a variable x was done using the command

```
shapito.test(log(x))
```

The Industry variable in the .csv file was used to carry out the industry grouping, using the *subset* command in R in the following fashion:

```
## Create the industry groups
C <- subset(fulldata, Industry=="C")
D <- subset(fulldata, Industry=="D")
G <- subset(fulldata, Industry=="G")
# etc
```

The response variables used as inputs in the overall test, and the bivariate equivalence test in Sections 3.4 and 3.5 were defined as follows:

```
## Calculate the overall mean for the Absolute and Growth emissions for the test in
    Section 3.4 and 3.5:
# The column numbers in the csv file for Absolute Emissions Year 1 was 9, for Year 2
     was 13, hence dividing column 13 by column 9 yielded the energy savings ratio
overallmeanabs <- c((sum(A[,13])/sum(A[,9])), (sum(C[,13])/sum(C[,9])), (sum(D[,13])
    /sum(D[,9])), (sum(G[,13])/sum(G[,9])), (sum(H[,13])/sum(H[,9])), (sum(I[,13])/
    sum(I[,9])), (sum(J[,13])/sum(J[,9])), (sum(K[,13])/sum(K[,9])), (sum(O[,13])/
    sum(O[,9])), (sum(P[,13])/sum(P[,9])))
# Similarly for the Growth response, using columns 12 and 17
overallmeangrowth <- c((sum(A[,17])/sum(A[,12])), (sum(C[,17])/sum(C[,12])), (sum(D
    [,17])/sum(D[,12])), (sum(G[,17])/sum(G[,12])), (sum(H[,17])/sum(H[,12])), (sum(
    I[,17])/sum(I[,12])), (sum(J[,17])/sum(J[,12])), (sum(K[,17])/sum(K[,12])), (sum
    (O[,17])/sum(O[,12])), (sum(P[,17])/sum(P[,12])))

## Create the GCI for bivariate equivalence test
bivartest <- BivariateComparison(overallmeanabs, overallmeangrowth, 5000, 0.05)

## Create the GCI and p-value for the overall significance test
overalltest <- CIFunction(overallmeanabs, 0, 5000, 0.05)
```

For Section 3.6, for the two-group test each industry group was further subdivided into those with AMR $= 0$ and AMR > 0, and aggregated in the same fashion:

```
# Create two groups for each industry classification, amr0 with no AMR and amr1 with
     some AMR
C_amr0 <- subset(fulldata, Industry=="C" & AMRYear1==0)
C_amr1 <- subset(fulldata, Industry=="C" & AMRYear1>0)
D_amr0 <- subset(fulldata, Industry=="D" & AMRYear1==0)
D_amr1 <- subset(fulldata, Industry=="D" & AMRYear1>0)
G_amr0 <- subset(fulldata, Industry=="G" & AMRYear1==0)
G_amr1 <- subset(fulldata, Industry=="G" & AMRYear1>0)
# etc
```

```
## Calculate the response variables to be compared, with column numbers 13 and 9
    corresponding to the Absolute Emissions variable
amr0 <- c((sum(A_amr0[,13])/sum(A_amr0[,9])), (sum(C_amr0[,13])/sum(C_amr0[,9])), (
    sum(D_amr0[,13])/sum(D_amr0[,9])), (sum(G_amr0[,13])/sum(G_amr0[,9])), (sum(H_
    amr0[,13])/sum(H_amr0[,9])), (sum(I_amr0[,13])/sum(I_amr0[,9])), (sum(J_amr0
    [,13])/sum(J_amr0[,9])), (sum(K_amr0[,13])/sum(K_amr0[,9])), (sum(O_amr0[,13])/
    sum(O_amr0[,9])), (sum(P_amr0[,13])/sum(P_amr0[,9])))
amr1 <- c((sum(A_amr1[,13])/sum(A_amr1[,9])), (sum(C_amr1[,13])/sum(C_amr1[,9])), (
    sum(D_amr1[,13])/sum(D_amr1[,9])), (sum(G_amr1[,13])/sum(G_amr1[,9])), (sum(H_
    amr1[,13])/sum(H_amr1[,9])), (sum(I_amr1[,13])/sum(I_amr1[,9])), (sum(J_amr1
    [,13])/sum(J_amr1[,9])), (sum(K_amr1[,13])/sum(K_amr1[,9])), (sum(O_amr1[,13])/
    sum(O_amr1[,9])), (sum(P_amr1[,13])/sum(P_amr1[,9])))

## Carry out the two means test
twoamrtest <- comparison(amr1,amr0, 5000, 0.05)
```

For the test in Section 3.6 using 4 groups, the same methodology was used:

```
# Create four groups for each industry group, amr0, amr1, amr2 and amr3 with
    different % of AMR coverage
C_amr0 <- subset(fulldata, Industry=="C" & AMRYear1==0)
C_amr1 <- subset(fulldata, Industry=="C" & AMRYear1>0 & AMRYear1<=33)
C_amr2 <- subset(fulldata, Industry=="C" & AMRYear1>33 & AMRYear1<=66)
C_amr3 <- subset(fulldata, Industry=="C" & AMRYear1>66)
D_amr0 <- subset(fulldata, Industry=="D" & AMRYear1==0)
D_amr1 <- subset(fulldata, Industry=="D" & AMRYear1>0 & AMRYear1<=33)
D_amr2 <- subset(fulldata, Industry=="D" & AMRYear1>33 & AMRYear1<=66)
D_amr3 <- subset(fulldata, Industry=="D" & AMRYear1>66)
G_amr0 <- subset(fulldata, Industry=="G" & AMRYear1==0)
G_amr1 <- subset(fulldata, Industry=="G" & AMRYear1>0 & AMRYear1<=33)
G_amr2 <- subset(fulldata, Industry=="G" & AMRYear1>33 & AMRYear1<=66)
G_amr3 <- subset(fulldata, Industry=="G" & AMRYear1>66)
# etc

## Construct four variables with the means from each group, to be used as the inputs
     for the simultaneous pairwise comparisons
amr0 <- c((sum(A_amr0[,13])/sum(A_amr0[,9])), (sum(C_amr0[,13])/sum(C_amr0[,9])), (
    sum(D_amr0[,13])/sum(D_amr0[,9])), (sum(G_amr0[,13])/sum(G_amr0[,9])), (sum(H_
    amr0[,13])/sum(H_amr0[,9])), (sum(I_amr0[,13])/sum(I_amr0[,9])), (sum(J_amr0
    [,13])/sum(J_amr0[,9])), (sum(K_amr0[,13])/sum(K_amr0[,9])), (sum(O_amr0[,13])/
    sum(O_amr0[,9])), (sum(P_amr0[,13])/sum(P_amr0[,9])))
# For amr1, C was removed as it had no responses fulfilling this criteria
```

```
amr1 <- c((sum(A_amr1[,13])/sum(A_amr1[,9])),  (sum(D_amr1[,13])/sum(D_amr1[,9])), (
    sum(G_amr1[,13])/sum(G_amr1[,9])), (sum(H_amr1[,13])/sum(H_amr1[,9])), (sum(I_
    amr1[,13])/sum(I_amr1[,9])), (sum(J_amr1[,13])/sum(J_amr1[,9])), (sum(K_amr1
    [,13])/sum(K_amr1[,9])), (sum(O_amr1[,13])/sum(O_amr1[,9])), (sum(P_amr1[,13])/
    sum(P_amr1[,9])))
amr2 <- c((sum(A_amr2[,13])/sum(A_amr2[,9])), (sum(C_amr2[,13])/sum(C_amr2[,9])), (
    sum(D_amr2[,13])/sum(D_amr2[,9])), (sum(G_amr2[,13])/sum(G_amr2[,9])), (sum(H_
    amr2[,13])/sum(H_amr2[,9])), (sum(I_amr2[,13])/sum(I_amr2[,9])), (sum(J_amr2
    [,13])/sum(J_amr2[,9])), (sum(K_amr2[,13])/sum(K_amr2[,9])), (sum(O_amr2[,13])/
    sum(O_amr2[,9])), (sum(P_amr2[,13])/sum(P_amr2[,9])))
# For amr3, M was removed for no responses, and H as it had only one response
amr3 <- c((sum(A_amr3[,13])/sum(A_amr3[,9])), (sum(C_amr3[,13])/sum(C_amr3[,9])), (
    sum(D_amr3[,13])/sum(D_amr3[,9])), (sum(G_amr3[,13])/sum(G_amr3[,9])),  (sum(I_
    amr3[,13])/sum(I_amr3[,9])), (sum(J_amr3[,13])/sum(J_amr3[,9])), (sum(K_amr3
    [,13])/sum(K_amr3[,9])), (sum(O_amr3[,13])/sum(O_amr3[,9])), (sum(P_amr3[,13])/
    sum(P_amr3[,9])))

## Simultaneous pairwise GCI's
data <- list(amr0, amr1, amr2, amr3)
fouramrtest <- severalmeans(data, 5000, 0.05)
```

In Section 3.7, the same procedure was used as for AMR, but with EMS. For the two group comparison test:

```
## Create two groups, ems0 and ems1
C_ems0 <- subset(fulldata, Industry=="C" & StandardYear1==0)
C_ems1 <- subset(fulldata, Industry=="C" & StandardYear1>0)
D_ems0 <- subset(fulldata, Industry=="D" & StandardYear1==0)
D_ems1 <- subset(fulldata, Industry=="D" & StandardYear1>0)
G_ems0 <- subset(fulldata, Industry=="G" & StandardYear1==0)
G_ems1 <- subset(fulldata, Industry=="G" & StandardYear1>0)
# etc

## Calculate the two response variables to be used in the "comparison" custom
    function
ems0 <- c((sum(A_ems0[,13])/sum(A_ems0[,9])), (sum(C_ems0[,13])/sum(C_ems0[,9])), (
    sum(D_ems0[,13])/sum(D_ems0[,9])), (sum(G_ems0[,13])/sum(G_ems0[,9])), (sum(H_
    ems0[,13])/sum(H_ems0[,9])), (sum(I_ems0[,13])/sum(I_ems0[,9])), (sum(J_ems0
    [,13])/sum(J_ems0[,9])), (sum(K_ems0[,13])/sum(K_ems0[,9])), (sum(O_ems0[,13])/
    sum(O_ems0[,9])), (sum(P_ems0[,13])/sum(P_ems0[,9])))
ems1 <- c((sum(A_ems1[,13])/sum(A_ems1[,9])), (sum(C_ems1[,13])/sum(C_ems1[,9])), (
    sum(D_ems1[,13])/sum(D_ems1[,9])), (sum(G_ems1[,13])/sum(G_ems1[,9])), (sum(H_
```

```
    ems1[,13])/sum(H_ems1[,9])), (sum(I_ems1[,13])/sum(I_ems1[,9])), (sum(J_ems1
    [,13])/sum(J_ems1[,9])), (sum(K_ems1[,13])/sum(K_ems1[,9])), (sum(O_ems1[,13])/
    sum(O_ems1[,9])), (sum(P_ems1[,13])/sum(P_ems1[,9])))

## Carry out the two means test
twoemstest <- comparison(ems1,ems0, 5000, 0.05)
```

Again, for Section 3.7 with four groups, the same method was used:

```
## Create four subgroups for different % of EMS, ems0, ems1, ems2 and ems3
C_ems0 <- subset(fulldata, Industry=="C" & StandardYear1==0)
C_ems1 <- subset(fulldata, Industry=="C" & StandardYear1>0 & StandardYear1<=33)
C_ems2 <- subset(fulldata, Industry=="C" & StandardYear1>33 & StandardYear1<=66)
C_ems3 <- subset(fulldata, Industry=="C" & StandardYear1>66)
D_ems0 <- subset(fulldata, Industry=="D" & StandardYear1==0)
D_ems1 <- subset(fulldata, Industry=="D" & StandardYear1>0 & StandardYear1<=33)
D_ems2 <- subset(fulldata, Industry=="D" & StandardYear1>33 & StandardYear1<=66)
D_ems3 <- subset(fulldata, Industry=="D" & StandardYear1>66)
G_ems0 <- subset(fulldata, Industry=="G" & StandardYear1==0)
G_ems1 <- subset(fulldata, Industry=="G" & StandardYear1>0 & StandardYear1<=33)
G_ems2 <- subset(fulldata, Industry=="G" & StandardYear1>33 & StandardYear1<=66)
G_ems3 <- subset(fulldata, Industry=="G" & StandardYear1>66)

## Calculate the response variables to be used:
ems0 <- c((sum(A_ems0[,13])/sum(A_ems0[,9])), (sum(C_ems0[,13])/sum(C_ems0[,9])), (
    sum(D_ems0[,13])/sum(D_ems0[,9])), (sum(G_ems0[,13])/sum(G_ems0[,9])), (sum(H_
    ems0[,13])/sum(H_ems0[,9])), (sum(I_ems0[,13])/sum(I_ems0[,9])), (sum(J_ems0
    [,13])/sum(J_ems0[,9])), (sum(K_ems0[,13])/sum(K_ems0[,9])), (sum(O_ems0[,13])/
    sum(O_ems0[,9])), (sum(P_ems0[,13])/sum(P_ems0[,9])))
# For ems1, we removed groups A, C, H, J & M as they had insufficient number of
    responses
ems1 <- c((sum(D_ems1[,13])/sum(D_ems1[,9])), (sum(G_ems1[,13])/sum(G_ems1[,9])), (
    sum(I_ems1[,13])/sum(I_ems1[,9])), (sum(K_ems1[,13])/sum(K_ems1[,9])), (sum(O_
    ems1[,13])/sum(O_ems1[,9])), (sum(P_ems1[,13])/sum(P_ems1[,9])))
# For ems2, we have taken out A, M, O
ems2 <- c((sum(C_ems2[,13])/sum(C_ems2[,9])), (sum(D_ems2[,13])/sum(D_ems2[,9])), (
    sum(G_ems2[,13])/sum(G_ems2[,9])), (sum(H_ems2[,13])/sum(H_ems2[,9])), (sum(I_
    ems2[,13])/sum(I_ems2[,9])), (sum(J_ems2[,13])/sum(J_ems2[,9])), (sum(K_ems2
    [,13])/sum(K_ems2[,9])), (sum(P_ems2[,13])/sum(P_ems2[,9])))
# For ems3, we have removed groups H and M
ems3 <- c((sum(A_ems3[,13])/sum(A_ems3[,9])), (sum(C_ems3[,13])/sum(C_ems3[,9])), (
    sum(D_ems3[,13])/sum(D_ems3[,9])), (sum(G_ems3[,13])/sum(G_ems3[,9])), (sum(I_
    ems3[,13])/sum(I_ems3[,9])), (sum(J_ems3[,13])/sum(J_ems3[,9])), (sum(K_ems3
```

```
    [,13])/sum(K_ems3[,9])), (sum(O_ems3[,13])/sum(O_ems3[,9])), (sum(P_ems3[,13])/
    sum(P_ems3[,9])))

## Pairwise comparisons among the 4 groups
data <- list(ems0, ems1, ems2, ems3)
fouremstest <- severalmeans(data, 5000, 0.1)
```

The GCI values and p-value can then be called with the appropriate function-name, see details of possible call options in B.2 above:

```
twoemstest$UpperCI
twoemstest$LowerCI
twoemstest$pvalue
```

Appendix C

Proof of Theorem 1

Proof of Theorem 1.

Proof. By the central limit theorem, we have

$$\sqrt{N}\left((\eta_{12}-\theta_{12}),(\eta_{13}-\theta_{13}),\cdots,(\eta_{k-1,k}-\theta_{k-1,k})\right) \xrightarrow{d} N(0,\mathbf{U}),$$

where $\mathbf{U}$ is an $k(k-1)/2 \times k(k-1)/2$ positive definite matrix. Let $u_{ab}, a, b = 1, 2, \cdots, k(k-1)/2$ be its (a,b)th entry. It can be shown that

$$u_{aa} = \frac{\sigma_i^2}{\lambda_i} + \frac{\sigma_i^4}{2\lambda_i} + \frac{\sigma_j^2}{\lambda_j} + \frac{\sigma_j^4}{2\lambda_j}$$

and

$$NV_{ij} \to \frac{\sigma_i^2}{\lambda_i} + \frac{\sigma_i^4}{2\lambda_i} + \frac{\sigma_j^2}{\lambda_j} + \frac{\sigma_j^4}{2\lambda_j}$$

almost surely. Therefore,

$$\left(\frac{\eta_{12}-\theta_{12}}{\sqrt{V_{12}}}, \frac{\eta_{13}-\theta_{13}}{\sqrt{V_{13}}}, \cdots, \frac{\eta_{k-1,k}-\theta_{k-1,k}}{\sqrt{V_{k-1,k}}}\right) \xrightarrow{d} N(0,\mathbf{U}^*),$$

where the (a,b)th entry of $\mathbf{U}^*$ is $u_{ab}/\sqrt{u_{aa}u_{bb}}$.

Take a random vector $(Z_1, Z_2, \cdots, Z_{k(k-1)/2})$ distributed according to $N(0,\mathbf{U}^*)$. By the continuous mapping theorem

$$\max_{i<j}\left|\frac{\theta_{ij}-\eta_{ij}}{\sqrt{V_{ij}}}\right| \xrightarrow{d} \max|Z_a|$$

for $1 \le a \le k(k-1)/2$.

For $i = 1, \cdots, k$, $U_i^2/n_i \xrightarrow{p} 1$. For all $i \neq j$,

$$\begin{aligned}
\sqrt{N}(R_{\theta_{ij}} - \eta_{ij}) &= \sqrt{N}\left\{-\sqrt{\frac{n_i-1}{n_i}}\cdot\frac{S_iZ_i}{U_i} + \sqrt{\frac{n_j-1}{n_j}}\cdot\frac{S_jZ_j}{U_j}\right.\\
&+ \left.\frac{(n_i-1)S_i^2}{2U_i^2} - \frac{(n_j-1)S_j^2}{2U_j^2} - \frac{n_i-1}{2(n_i-3)}S_i^2 + \frac{n_j-1}{2(n_j-3)}S_j^2\right\}\\
&= \frac{\sigma_i}{\sqrt{\lambda_i}}Z_j - \frac{\sigma_i}{\sqrt{\lambda_i}}Z_i + o_p(1) \qquad \text{(C.1)}
\end{aligned}$$

conditionally on $T = (\bar{\mathbf{Y}}, \mathbf{S}^2)$ almost surely.

Recall that $NV_{ij} \to \frac{\sigma_i^2}{\lambda_i} + \frac{\sigma_i^4}{2\lambda_i} + \frac{\sigma_j^2}{\lambda_j} + \frac{\sigma_j^4}{2\lambda_j}$ almost surely and note that

$$\begin{aligned}
NR_{\xi_{ij}} &= N\frac{(n_i-1)S_i^2}{n_iU_i^2} + N\frac{(n_i-1)^2}{2n_i(n_i-3)^2}\left(\frac{(n_i-1)S_i^2}{U_i^2}\right)^2\\
&+ N\frac{(n_j-1)S_i^2}{n_jU_j^2} + N\frac{(n_j-1)^2}{2n_j(n_j-3)^2}\left(\frac{(n_i-1)S_i^2}{U_i^2}\right)^2\\
&= \frac{\sigma_i^2}{\lambda_i} + \frac{\sigma_i^4}{2\lambda_i} + \frac{\sigma_j^2}{\lambda_j} + \frac{\sigma_j^4}{\lambda_j} + o_p(1)
\end{aligned}$$

conditionally on T almost surely. It can be shown that Equation (C.1) implies

$$\max_{i<j}\left|\frac{\theta_{ij}-\eta_{ij}}{\sqrt{R_{\xi_{ij}}}}\right| \xrightarrow{d} \max_{1\le a\le k(k-1)/2}|Z_a| \qquad \text{(C.2)}$$

on T almost surely. Let F be the cdf of $\max_{1\le a\le k(k-1)/2}|Z_a|$. By the continuity of F

$$\sup_x|F_n(x|T) - F(x)| \to 0$$

almost surely, where F_n is the conditional distribution function of the left side of

(C.2). As a result,

$$
\begin{aligned}
P\left(\theta_{ij} \in \eta_{ij} \pm q(\alpha)\sqrt{V_{ij}} \quad \text{for all} \quad i < j\right) &= P\left\{F_n\left(\max_{i<j}\left|\frac{\theta_{ij}-\eta_{ij}}{\sqrt{V_{ij}}}\right|\middle| T\right) \leq 1-\alpha\right\} \\
&= P\left\{F\left(\max_{i<j}\left|\frac{\theta_{ij}-\eta_{ij}}{\sqrt{R_{\xi_{ij}}}}\right|\right) + o_p(1) \leq 1-\alpha\right\} \\
&\xrightarrow{d} 1-\alpha
\end{aligned}
$$

□

References

Alexander, R., & Govern, D. M. (1994). A new and simpler approximation and ANOVA under variance heterogeneity. *Jounal of Educational Statistics*, *19*, 91–101.

Angus, J. E., & Angust, B. J. E. (1994). Bootstrap one-sided confidence intervals for the log-normal mean. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *43*(3), 395–401.

Antoniou, I., Ivanov, V., Ivanov, V., & Zrelov, P. (2004, January). On the log-normal distribution of stock market data. *Physica A: Statistical Mechanics and its Applications*, *331*(3-4), 617–638. doi: 10.1016/j.physa.2003.09.034

Antoniou, I., Ivanov, V., Ivanov, V. V., & Zrelov, P. (2002, July). On the log-normal distribution of network traffic. *Physica D: Nonlinear Phenomena*, *167*(1-2), 72–85. doi: 10.1016/S0167-2789(02)00431-1

Arimura, T., Li, S., Newell, R., & Palmer, K. (2011, October). Cost-Effectiveness of Electricity Energy Efficiency Programs. *Discussion Paper, Resources for the Future*. Retrieved from `http://www.nber.org/papers/w17556.pdf` doi: 10.3386/w17556

Bebu, I., & Mathew, T. (2008). Comparing the means and variances of a bivariate log-normal distribution. *Statistics in Medicine*(27), 2684–2696. doi: 10.1002/sim

Bundesamt fur Statistik. (1997). *Statistisches Jahrbuch der Schweiz.* Verlag Neue Zricher Zeitung.

Carbon Trust. (2014). *Carbon Trust Standard.* `http://www.carbontrust.com/client-services/footprinting/footprint-certification/carbon-trust-standard`.

Casella, G., & Berger, R. L. (2002). *Statistical Inference, 2nd Edition.* Duxbury Thomson Learning.

Daniel, J. W. (2008). *LogNormal stock-price models in Exams MFE / 3 and C / 4.* `www.actuarialseminars.com/Misc/Lognormal.pdf`.

Department of Energy and Climate Change. (2014). *CRC Energy Efficiency Scheme.* Retrieved from `https://www.gov.uk/government/policies/reducing-demand-for-energy-from-industry-businesses-and-the-public-sector--2/supporting-pages/crc-energy-efficiency-scheme`

Environment Agency. (2011). *2010/2011 - CRC Performance League Table.* Retrieved from `http://crc.environment-agency.gov.uk/pplt/web/plt/public/2010-11/CRCPerformanceLeagueTable20102011`

Environment Agency. (2012). *2011/2012 - CRC Performance League Table.* Retrieved from `http://crc.environment-agency.gov.uk/pplt/web/plt/public/2011-12/CRCPerformanceLeagueTable20112012`

Ferreira, V., Alves, L., Fleming, P., Stuart, G., Patel, P., Webber, P., & Conway, S. (2007). Low hanging fruits or cost-effective energy and water savings using intelligent metering and monitoring systems? *ECEEE Summer Study*, 739–742.

Guo, J., & Luh, W. (2000). Testing methods for the one-way fixed effects ANOVA models of log-normal samples. *Jounal of Applied Statistics*, *27*, 731–738.

Gupta, R., & Li, X. (2006). Statisitcal inference for the common mean of two log-normal samples. *Computational Statistics and Data Analysis*, *50*, 3141–3164.

Hannig, J., Iyer, H., & Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of American Statistical Association*, *101*, 254–269.

Hannig, J., Lidong, Abdel-Karim, A., & Iyer, H. (2006). Simultaneous fiducial generalized confidence intervals for ratios of means of lognormal distributions. *Austrian Journal of Statistics*, *35*, 261–269.

Jaffe, A., & Stavins, R. (1995). Dynamic Incentives of Environmental Regulations: The Effects of Alternative Policy Instruments on Technology Diffusion. *Journal of Environmental Economics and Management*, *29(3)*, 43-63.

James, G. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, *38*, 324–329.

Kondo, K. (1977). The log-normal distribution of the incubation time of exogenous diseases. *Japanese Journal of Human Genetics*, *21*, 217–237.

Krishnamoorthy, K., & Mathew, T. (2003). Inferences on the means of lognormal distributions using generalized p -values and generalized confidence intervals. *Journal of Statistical Planning and Inference*, *115*, 103–121.

Land, E. (1973). Standard Confidence for Linear Functions Limits of the Normal Mean and Variance. *Journal of the American Statistical Association*, *68*(344), 960–963.

Li, X. (2009, June). A generalized -value approach for comparing the means of several log-normal populations. *Statistics & Probability Letters*, *79*(11), 1404–1408. doi: 10.1016/j.spl.2009.03.004

Lien, D., & Balakrishnan, N. (2006, April). Moments and properties of multiplicatively constrained bivariate lognormal distribution with applications to futures hedging. *Journal of Statistical Planning and Inference*, *136*(4), 1349–1359. doi: 10.1016/j.jspi.2004.10.004

Magurran, A. E. (1998). *Ecological Diversity and Its Measurement.* Croom Helm.

Malanca, A. (1996). Distribution of 226Ra, 232Th, and 40K in soils of Rio Grande do Norte (Brazil). *Journal of Environmental Radioactivity*, *30*(1), 55–67.

Sharpe, M. J. (2004). *Lognormal model for stock prices.* `math.ucsd.edu/~msharpe/stockgrowth.pdf/`.

Tsui, K., & Weerahandi, S. (1989). Generalized pValues in Significance Testing of Hypotheses in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, *84*(406), 602–607.

Weerahandi, S. (1993). Generalized Confidence Intervals. *Journal of the American Statistical Association*, *88*(423), 899–905.

Welch, B. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, *38*, 330–336.

Williams, C. B. (1940). A Note on the Statistical Analysis of Sentence-Length as a Criterion of Literary Style. *Biometrica*, *31*(3/4), 356–361.

Xiong, S., & Mu, W. (2009). Simultaneous confidence intervals for for one-way layout based on generalized pivotal quantities. *Journal of Statistical Computing and Simulation*, *79*, 1235-1244.

Zhou, X., & Gao, S. (1997). Confidence intervals for the lognormal mean. *Statistics in Medicine*, *16*(7), 783–790.

Zhou, X., Gao, S., & Hui, S. (1997). Methods for comparing means of two independent log-normal samples. *Biometrics*, *53*, 1129–1135.