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

Bose Falk ¹ and Guoyi Zhang ²

Abstract

Generalized Confidence Intervals (GCI) can be constructed for cases where an exact confidence interval based on sufficient statistics is not available. In this article, we propose fiducial generalized pivotal quantities (FGPQ)-based simultaneous confidence intervals for ratios of log-normal means. We then apply the proposed method and several existing GCI approaches to a dataset from Carbon Reduction Commitment Energy Efficiency Scheme (CRC) to test energy saving percentages among different groups.

Key Words: Carbon Reduction Commitment Energy Efficiency Scheme, Energy Saving, Fiducial Generalized Pivotal Quantities, Generalized Confidence Intervals, Multiple Comparison.

1 Introduction

The CRC is a mandatory scheme aimed at improving energy efficiency and cutting emissions in large public and private sector organizations in the United Kingdom (UK) (Department of Energy and Climate Change, 2014). This government scheme requires all large organizations that meet the specified participation criteria to report their total energy use every year and pay a levy for each tCO₂ (ton of carbon dioxide or equivalent gasses) emitted. Many organizations also report their energy intensity (emissions over revenue) each year, commonly referred to as the “Growth Emissions”. In the first year, the organizations also report their

¹Bose Falk, Graduate student, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, 87131-0001

²Guoyi Zhang: Corresponding author, Assistant Professor, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, 87131-0001 (gzhang123@gmail.com)
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. The CRC data has attracted a lot of attentions in recent years. Some standard methods, however, do not apply to the dataset directly, as some of the variables do not follow a normal distribution even after transformation. GCI approaches are therefore adopted in analyzing the data. In order to find out if there are AMR and EMS effects on energy savings, we grouped the observations into representative groups according to AMR and EMS levels. We found out that after grouping, the variables are log-normally distributed. To perform a pairwise comparison among the different groups, we require a new method in multiple comparison for several log-normal distributions.

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. Simultaneous confidence intervals for certain log-normal parameters are useful in many areas. For example, in pharmaceutical statistics, it is often of interest to compare the mean responses of several drugs to ensure that they are (more or less) equally effective. In standard analysis of variance, Scheffé’s method (Scheffé, 1959), the Bonferroni inequality-based method, and Tukey method (Tukey, 1953) are widely used for simultaneous pairwise comparisons. When variances are heteroscedastic and group sizes are unequal, exact frequentist tests are unavailable. In such situations, parametric bootstrap and generalized p-value (Tsui & Weerahandi, 1989) procedures are commonly used. Weerahandi (1993) introduced the concept of a generalized pivotal quantity. Later, Hannig, Iyer, and Patterson (2006) introduced a subclass of Weerahandi’s generalized pivotal quantities, called FGPQs, and provided procedures to derive FGPQs. Hannig, E, Abdel-Karim, and Iyer (2006) also proposed simultaneous fiducial generalized confidence intervals 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, E, 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, E, et al. (2006)’s methods. Follow Xiong and Mu (2009)’s idea, we proposed FGPQ-based MCP for ratios of means from several log-normal populations under heteroscedasticity.

This article is outlined as follows. In Section 2, we describe the CRC dataset. In Section 3, we propose fiducial generalized pivotal quantities (FGPQ)-based simultaneous confidence intervals for ratios of log-normal means. In Section 4, we report data analysis results. In Section 5, we give conclusions and propose future work.

2 Dataset

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. The energy consumption in MWh reported by qualifying organizations is used to calculate tCO$_2$ based on the UK’s energy mix for each year. 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 our data set is based on. We are interested in the following variables from Year 1 (Financial Year 2010 / 2011) to Year 2 (Financial Year 2011 / 2012):

- **Absolute Emissions**: the absolute emissions in tCO$_2$ is calculated by the environment agency by converting the reported energy consumption in MWh to tCO$_2$ using each year’s Energy Mix. The savings ratio are the current year’s emissions divided by the rolling historical average emissions;

- **Growth Emissions**: 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;

- **The Standard Industry (SIC) Code**: usually in the following format: N.23.12, where N denotes the overall industry, the first number group (23) denotes 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. Around 34% of the organizations had no AMR in Year 1. This variable is only reported in the first two years;

- EMS: the percentage of emissions in that year covered by an Energy Management Scheme (EMS). Around 65% of the organizations had no EMS in Year 1, and it is only reported in the first two years.

We use data for organizations that reported their absolute and optional growth emissions in both 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). The total number of useable observations is 1314. Organizations with AMR and EMS coverage were given higher positioning in the public PLT for the first two years of the scheme (when no historical data was available), since these two measures indicated that an organization was actively working to improve their energy efficiency. We are therefore interested in finding out whether organizations with 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,
\[
y = \frac{\text{Emissions Year 2}}{\text{Emissions Year 1}}.
\]
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. Our main interests in testing is as follows:

**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 four 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 levels.

3 Methodology

The idea of GCI (Weerahandi, 1993) is to construct confidence intervals for cases where exact frequentist tests based on sufficient statistics are not available. Krishnamoorthy and Mathew (2003) proposed GCI for testing a log-normal mean and comparing two independent log-normal means. Bebu and Mathew (2008) proposed GCI for testing equivalence of bivariate log-normal means. In this section, we propose FGPQ-based MCP for ratios of means from several log-normal populations under heteroscedasticity (for more details, see Zhang and Falk (2014)). First, let’s give a background review for generalized variable approach.

3.1 Generalized variable approach

Let \( Y_{ij}, i = 1, \cdots, k, j = 1, \cdots, n_i \) be a random sample from \( k \) log-normal distributions with parameters \( \mu_i \) and \( \sigma_i^2 \), and let \( X_{ij} = \log Y_{ij} \). By definition, \( X_{ij}, j = 1, \cdots, n_i \) is an independent random sample from the \( k \) populations and has a normal distribution of \( N(\mu_i, \sigma_i^2) \). For each sample, the sample mean and variance are defined as follows

\[
\bar{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}, S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2.
\]

Let

\[
Z_i = \sqrt{n_i}(\bar{X}_i - \mu_i)/\sigma_i \quad \text{and} \quad U_i^2 = (n_i - 1)S_i^2/\sigma_i^2.
\]

It is well known that \( Z_i \sim N(0, 1) \) and \( U_i^2 \sim \chi^2_{(n_i-1)} \) and they are jointly independent. For each population, define

\[
M_i = E(Y_{ij}) = e^{\mu_i + \sigma_i^2/2} \quad \text{and} \quad \theta_i = \log(M_i) = \mu_i + \sigma_i^2/2. \quad (1)
\]
Krishnamoorthy and Mathew (2003) suggested the following generalized pivotal variables for $\mu_i$ and $\sigma^2_i$:

$$T_{\mu_i} = \bar{x}_i - \bar{X}_i - \frac{Z_i}{U_i} \cdot \frac{s_i}{\sqrt{n_i}}\sqrt{n_i} = \bar{x}_i - \frac{n_i - 1}{n_i} \cdot \frac{Z_i s_i}{U_i},$$  \hspace{1cm} (2)$$

and

$$T_{\sigma^2_i} = \frac{s^2_i}{S^2_i} = \frac{s^2_i}{U^2_i/(n_i - 1)},$$  \hspace{1cm} (3)$$

where $\bar{x}_i$ and $s^2_i$ are the observed values of $\bar{X}_i$ and $S^2_i$.

### 3.2 GCI for simultaneous pairwise comparison

The multiple comparison testing problem is as follows:

$$H_0 : M_i = M_j \text{ for all } i \neq j \quad \text{vs} \quad H_\alpha : \text{at least one } M_i \neq M_j.$$  \hspace{1cm} (4)$$

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

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

The problem of constructing simultaneous confidence intervals for $M_{ij}$ is equivalent to the problem of constructing simultaneous confidence intervals for $\theta_{ij}$. The multiple comparison problem in (4) is equivalent to the hypothesis tests

$$H_0 : \text{All } \theta_{ij} = 0 \quad \text{vs} \quad H_\alpha : \text{Not all } \theta_{ij} = 0.$$  \hspace{1cm} (5)$$

Follow Xiong and Mu (2009), we can define the FGPOs for $\mu_i$ and $\sigma^2_i$ (for $i = 1, ..., k$) as $R_{\mu_i}$ and $R_{\sigma^2_i}$:

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

$$R_{\sigma^2_i} = \frac{(n_i - 1)S^2_i}{U^2_i},$$

which then gives us the pivotal statistic for $\theta_i$, $R_{\theta_i}$:

$$R_{\theta_i} = R_{\mu_i} + \frac{R_{\sigma^2_i}}{2} = \bar{X}_i - \frac{n_i - 1}{n_i} \cdot \frac{S^2_i Z_i}{U_i} + \frac{(n_i - 1)S^2_i}{2U^2_i},$$

$$6$$
As a result, the FGPQ for \( \theta_{ij} \) is
\[
R_{\theta_{ij}} = \bar{X}_i - \bar{X}_j - \frac{n_i - 1}{n_i} S_i Z_i + \frac{n_j - 1}{n_j} S_j Z_j + \frac{(n_i - 1) S_i^2}{2 U_i^2} - \frac{(n_j - 1) S_j^2}{2 U_j^2}. \tag{5}
\]

Let \( \bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k), S^2 = (S_1^2, S_2^2, \ldots, S_k^2) \). The conditional expectation of \( \eta \) and variance of \( R_{\theta_{ij}} \) can be derived as
\[
\eta_{ij} = E(R_{\theta_{ij}} | \bar{X}, S^2) = \bar{X}_i - \bar{X}_j + \frac{n_i - 1}{2(n_i - 3)} S_i^2 - \frac{n_i - 1}{2(n_i - 3)} S_i^2, \tag{6}
\]
\[
V_{ij} = Var(R_{\theta_{ij}} | \bar{X}, S^2) = \frac{n_i - 1}{n_i(n_i - 3)} S_i^2 + \frac{(n_i - 1)^2}{2(n_i - 3)(n_i - 5)} S_i^4 + \frac{n_j - 1}{n_j(n_j - 3)} S_j^2 + \frac{(n_j - 1)^2}{2(n_j - 3)(n_j - 5)} S_j^4. \tag{7}
\]

If \( \xi_{ij} \) is the variance of \( \eta_{ij} \), and \( R_{\xi_{ij}} \) the pivotal statistic of \( \xi_{ij} \), then:
\[
\xi_{ij} = \text{Var}\{E(R_{\theta_{ij}} | \bar{X}, S^2)\}
= \frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j} + \left( \frac{n_i - 1}{2(n_i - 3)} \right)^2 \frac{2 \sigma_i^4}{n_i} + \left( \frac{n_j - 1}{2(n_j - 3)} \right)^2 \frac{2 \sigma_j^4}{n_j} \tag{8}
= \frac{\sigma_i^2}{n_i} + \frac{(n_i - 1)^2}{2n_i(n_i - 3)^2} + \frac{\sigma_j^2}{n_j} + \frac{(n_j - 1)^2}{2n_j(n_j - 3)^2},
\]
\[
R_{\xi_{ij}} = \frac{(n_i - 1) S_i^2}{n_i U_i^2} + \frac{(n_i - 1)^2}{2n_i(n_i - 3)^2} \left( \frac{(n_i - 1) S_i^2}{U_i^2} \right)^2 + \frac{(n_j - 1) S_j^2}{n_j U_j^2} + \frac{(n_j - 1)^2}{2n_j(n_j - 3)^2} \left( \frac{(n_j - 1) S_j^2}{U_j^2} \right)^2. \tag{9}
\]

As pointed by Xiong and Mu (2009), FGPQs can be used to provide effective approximations of distributions. The distribution of
\[
\max_{i \leq j} \left| \theta_{ij} - E(R_{\theta_{ij}} | \bar{X}, S^2) \right| \bigg/ \sqrt{\text{Var}(R_{\theta_{ij}} | \bar{X}, S^2)}, \tag{10}
\]
with
\[
Q = \max_{i \leq j} \left| \frac{R_{\theta_{ij}} - E(R_{\theta_{ij}} | \bar{X}, S^2)}{\sqrt{R_{\xi_{ij}}}} \right|. \tag{11}
\]
Then, the simultaneous confidence intervals for $\theta_{ij}$ is

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

where $q(\alpha)$ is the upper $\alpha$ quantile of $Q$ calculated by the following algorithm:

**Algorithm 1:**

1. For given observations $y_{ij}$ (where $i = 1, \ldots, k$, $j = 1, \ldots, n_i$) compute $x_{ij} = \log(y_{ij})$
2. Compute $\bar{x}_i$ 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. Compute $q(\alpha)$, the $(1 - \alpha)100\%$ percentile of $Q$.

Zhang and Falk (2014) proved that the constructed confidence intervals have correct asymptotic coverage. And their simulation studies show that the proposed methods work well in terms of coverage probabilities.

## 4 Data Analysis

### 4.1 Log-Normal Assumption and Grouping

The data set is downloadable from the Environment Agency’s website. It contains a variable for the percentage savings between Year 2 and the historical average (Year 1), for both Absolute and Growth Emissions. However, neither variable is normally distributed even after transformation.

To reduce variability, we group the organizations together according to industry groups by using the first letter in SIC code. 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). Among these 13 groups, there are five groups with fewer than 10 organizations. We group the organizations from these five groups into one industry group named as “Other”. This ensures that all industry groups have >10 organizations each. We therefore have 10 groups in total, 8 from SIC classification, one for public organizations, and

Figure 1: Boxplot of $\log(Y_{abs})$ for each industry grouping
after industry grouping gives the following results:

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

Hence after grouping by industry, our basic assumption of log-normality holds. 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 model as the data fulfills the three properties: positive only values, increased skewness and multiplicative property, where a log-normal model is widely used. At the time of writing this article, only two years of energy usage data is available. The CRC scheme is intended to continue for some time. If the data is 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.

### 4.2 Overall Significance of Savings Ratios

Before we proceed to test for saving difference between Year 1 and Year 2, we first want to decide a suitable response variable to evaluate the savings. There are two possible interested variables, the ratio of Absolute Emissions \( Y_{\text{abs}} = \text{Absolute Emissions}_2 / \text{Absolute Emissions}_1 \) and the ratio of Growth Emissions \( Y_{\text{growth}} = \text{Growth Emissions}_2 / \text{Growth Emissions}_1 \), that we can use to evaluate the overall savings. Let \( \eta_{\text{abs}} = \log E(Y_{\text{abs}}) \) and \( \eta_{\text{growth}} = \log E(Y_{\text{growth}}) \). We want to test:

\[ H_0 : \eta_{\text{abs}} = \eta_{\text{growth}} \quad \text{vs} \quad H_\alpha : \eta_{\text{abs}} \neq \eta_{\text{growth}} \]  \hspace{1cm} (13)

Using Bebu and Mathew (2008)’s test on equivalence of bivariate log-normal means, we wish to construct a confidence interval for \( \theta = \eta_{\text{abs}} - \eta_{\text{growth}} = (\mu_{\text{abs}} - \mu_{\text{growth}}) + \frac{1}{2}(\sigma_{\text{abs}}^2 - \sigma_{\text{growth}}^2) \). The GCI for \( \theta \) 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 that the ratio of the means of Absolute and Growth Emissions are not significantly different. Furthermore, Growth variable will no longer be reported in Year 3, and we observe that variance of the Growth variable is very high. Hence in order to make it easy to incorporate future data into this analysis, we proceed by using Absolute Emissions variable.

Now using the ratio of Absolute Emissions as the response variable, we want to investigate whether the savings in Year 2 are statistically significant from Year 1, or whether energy savings ratio $Y_{abs2}/Y_{abs1}$ is lower than the no change scenario. Since $Y_{abs}$ follows log-normal distribution, we can use the methodology from Krishnamoorthy and Mathew (2003) to test the following hypothesis:

\[ H_0 : \eta_{abs} \geq 0 \quad \text{vs} \quad H_\alpha : \eta_{abs} < 0. \]  

A 95% upper confidence interval for $\eta_{abs}$ of -0.0413 and a p-value for the null hypothesis in Eq (14) of $p = 0$. Hence, 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 will further continue our analysis with organizations by various AMR and EMS levels.

### 4.3 Automated Meter Reader

In the first year of the scheme, participants were asked to submit information on the percentage of their total emissions covered by AMR. AMR refers to technologies that automatically report energy usage data to a central location without the need for manual meter readings. AMR technologies also allow an organization to identify areas where energy usage could be streamlined, and get a real-time picture of the current energy usage. Employing AMR usually allows organizations to monitor their energy usage more efficiently, and hence more easily to identify saving options. AMR takes a value between 0 - 100%, with approximately one third of the reporting organizations having no AMR, i.e., $AMR = 0$. 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.

We first carry out a comparison between 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. To construct the two groups $G_1$ and $G_2$, we first subdivide each industry group into organizations with no AMR, and with some AMR. For example, for the organizations where Industry = C, we divide them into those with no AMR $G_{1C}$ and those with some AMR $G_{2C}$, and the mean absolute savings ratio of these organizations respectively is the mean of $G_{1C}$ and $G_{2C}$. Therefore $\eta_{G_1}$ is the log-normal mean of $G_{1j}$, where $j$ are the 10 industry groupings. We want to test

$$H_0 : \eta_{G_2} \geq \eta_{G_1} \quad \text{vs} \quad H_\alpha : \eta_{G_2} < \eta_{G_1}.$$  \hspace{1cm} (15)

Using Krishnamoorthy and Mathew (2003)'s methods, an upper level confidence interval value is 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 in Eq (15), and hence we conclude that organizations with some AMR did not achieve a better savings ratio compared to those with no AMR. Therefore, we proceed further by dividing the organizations into four segments according to AMR levels as follows:

- $A_1$: AMR % = 0
- $A_2$: 0 < AMR % ≤ 33%
- $A_3$: 33% < AMR % ≤ 66%
- $A_4$: 66% < AMR % ≤ 100%.

Note that group $G_1$ and $A_1$ both denote the organizations without AMR. We use $A_1$ for convenience of the subsequent notation of $A_2$ to $A_4$. For group $A_2$, there are no industries from group Industry = C that satisfies this criteria; For $A_4$, only one company from industry group H satisfies this criteria, therefore, this observation was removed to reduce variability.
Table 1: Multiple comparisons among four groups with different AMR levels

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.963</td>
<td>0.047</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-0.846</td>
<td>0.052</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-0.131</td>
<td>-0.006</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-0.058</td>
<td>0.074</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-0.106</td>
<td>0.017</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-0.111</td>
<td>0.006</td>
</tr>
</tbody>
</table>

All the four groups pass the Shapiro-Wilk test for log-normality at \( \alpha = 0.05 \). Thus, a multiple comparison test is as follows,

\[
H_0 : \eta_i - \eta_j = 0 \quad \text{vs} \quad H_\alpha : \eta_i - \eta_j \neq 0 \quad (16)
\]

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

Using our proposed methods in Section 3.2, with number of iterations \( m = 5000 \) from Algorithm 1. At the significance level \( \alpha = 0.05 \), we have listed the results in Table 1. From Table 1, we see that the only significant difference appears between groups \( A_1 \) and \( A_4 \). Organizations with an AMR percentage > 66% had a higher ratio than those with no AMR. We haven’t seen significant difference among other groups.

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 a more evenly matched starting points and the early adopters of energy efficiency aren’t “penalized” as described above.

4.4 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 (group $F_1$) and those with some EMS (group $F_2$). 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. We feel comfortable to assume log normality of the data. The following hypothesis is considered,

\[ H_0 : \eta_{F_2} \geq \eta_{F_1} \text{ vs } H_\alpha : \eta_{F_2} < \eta_{F_1} \]

(17)

Using the methods suggested by Krishnamoorthy and Mathew (2003), the upper level confidence interval is 0.0935, with a p-value of 0.944. Hence we cannot reject the null hypothesis, and conclude that, 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 following four groups according to EMS levels.

- $E_1$: EMS % = 0
- $E_2$: 0 < EMS % $\leq$ 33%
Table 2: Multiple comparisons among four groups with different EMS levels

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>Lower CI</th>
<th>Upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.051</td>
<td>0.112</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-0.127</td>
<td>0.089</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-0.098</td>
<td>0.011</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-0.177</td>
<td>0.077</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-0.163</td>
<td>0.014</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-0.137</td>
<td>0.088</td>
</tr>
</tbody>
</table>

- $E_3$: 33% < EMS % ≤ 66%
- $E_4$: 66% < EMS % ≤ 100%

All groups but $E_3$ passes the Shapiro-Wilk test of log-normality, but as the additional variance in $E_3$ is due to one single outlier (from Industry = P), we will continue and analyze all pairwise comparisons using the log-normal model. The multiple comparisons we consider is as follows,

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

where $i, j = 1, 2, 3, 4$, and also $i \neq j$. Using the proposed methods and $m = 5000$ iterations, at significance level $\alpha = 0.05$, the pairwise comparisons are listed in Table 2.

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.
5 Conclusion

Log-normal distribution is widely used to describe the distribution of positive random variables that exhibit skewness in biological, medical, economical and social studies. In this research, we proposed simultaneous confidence intervals for ratios of the means from \( k \) log-normally distributed data under heteroscedasticity.

Several 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 in general, with the exception that organizations with high AMR performed worse than organizations with no AMR. Organizations that had 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 that 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. Our research 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. 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.

References


