John Field Consulting Pty Ltd
ABN 62 096 615 235

10 High St, Burnside SA 5066
Telephone: (08) 8332 5294
Facsimile: (08) 8332 1229
Mobile: 0409 097 586
Email: JohnField@ozemail.com.au

Distribution of SAIDI data
Part 2

Report produced for

United Energy
Level 3, 501 Blackburn Road
MOUNT WAVERLEY Vic. 3149

Version 2, 26 October 2009

Distribution of SAIDI data
Part 2

Summary

This report examines extra SAIDI data in a bid to find a better transformation of SAIDI data for the Australian Energy Regulator's Service Target Performance Incentive Scheme. The following points emerge.

- The use of annual customer numbers instead of monthly ones makes little difference to SAIDI values for this analysis.
- There is a consistent pattern of non-normal behaviour of annual log(SAIDI) data, and in rolling 5-year periods of log(SAIDI).
- An examination of annual data from 1999 to 2008 suggests that the distribution of SAIDI has shifted upwards (ie towards higher SAIDI values) since 2003 – 2005.
- SAIDI values from both rural and urban feeders were examined separately for the two calendar years 2004-2005. There are distinct differences between rural and urban distributions of SAIDI, however consideration of the difference does not appear to help in transforming the overall distribution to normality.
- The main conclusion from the previous report still stands: it does not appear that a better transformation for SAIDI than log(SAIDI) can be found.

Table of Contents

1. Data

Several data sets were supplied for this analysis

DATA1: The data used for the previous report.
`Unplanned Daily Exclusion threshold calUED cl20091006.xls`, sheet `Unpl All`

DATA2: `1999-2005 daily SAIDI.xls`, sheet `SummaryData`

DATA3: Customer numbers by email 20/10/09, sourced from Craig Savage

DATA4: `2005DMSsummaryv2JTR.xls`, sheet `general`

DATA5: `rough ln charts SAIDI and SAIFIapr05JTR.xls`, sheets `Urban daily data` and `Rural daily data`

DATA6: `CustomernumbersdifferentsourcesJTR23102009.xls`, which includes DATA3.

The contents of these data sources is as follows

Source	Dates	SAIDI/CMOS		Customers	
		Which?	Urb/Rur split?	Basis	Urb/Rur split?
DATA1	1/01/04 – 31/08/09	Both	No	Monthly	No
DATA2	1/01/99 – 11/12/03 1/01/04 – 10/02/05	Both Both	No No	Annual Monthly	No No
DATA3	1999 – 2002 2003 – 2008	- -	- -	Annual Annual	No Yes
DATA4	1/01/05 -31/12/05	Both	Yes	Monthly, based on annual ratio	Yes
DATA5	1/01/99 – 31/12/04 excl 11-31/12/03	CMOS	Yes	-	-
DATA6	As for DATA3, plus extra information (not used)				

2. Effect of customer numbers

We can use the 2004 year to examine the effect of using yearly or monthly customer numbers. The monthly customer numbers from DATA2 are given below.

Jan	600,747	Jul	602,612
Feb	600,764	Aug	603,285
Mar	601,364	Sep	603,729
Apr	601,634	Oct	603,954
May	601,634	Nov	603,924
Jun	602,107	Dec	604,422

The weighted average of these is 602,521. This compares with 602,585 from DATA3. Thus the DATA3 customer figure appears close to the average of the DATA2 customer numbers.

The individual SAIDI values calculated from each data set agree to within ±0.3%.

A formal test of significance that the two distributions are different (a Kolmogorov-Smirnov test) showed no significant difference (P=1.00).

We may safely conclude that the use of yearly customer numbers instead of monthly ones will not adversely affect our investigation.

3. SAIDI distribution

This data set is available over a longer run than that previously used. Formal tests of normality all produce a significance probability of P < 2 x 10-16 for the entire period 1999 – 2008, showing that the data for the entire period cannot be regarded as being normally distributed..

A normal probability plot for each complete year is shown below in Figure 1. The general shape is the same for each year, although the lower tail for 1999 appears slightly different than the other years. The positions of the curves are a bit difficult to see, but in general terms they start from the top of the band (1999), reach the bottom in about 2003 – 2005, depending on the area of the distribution, and then work their way up again.

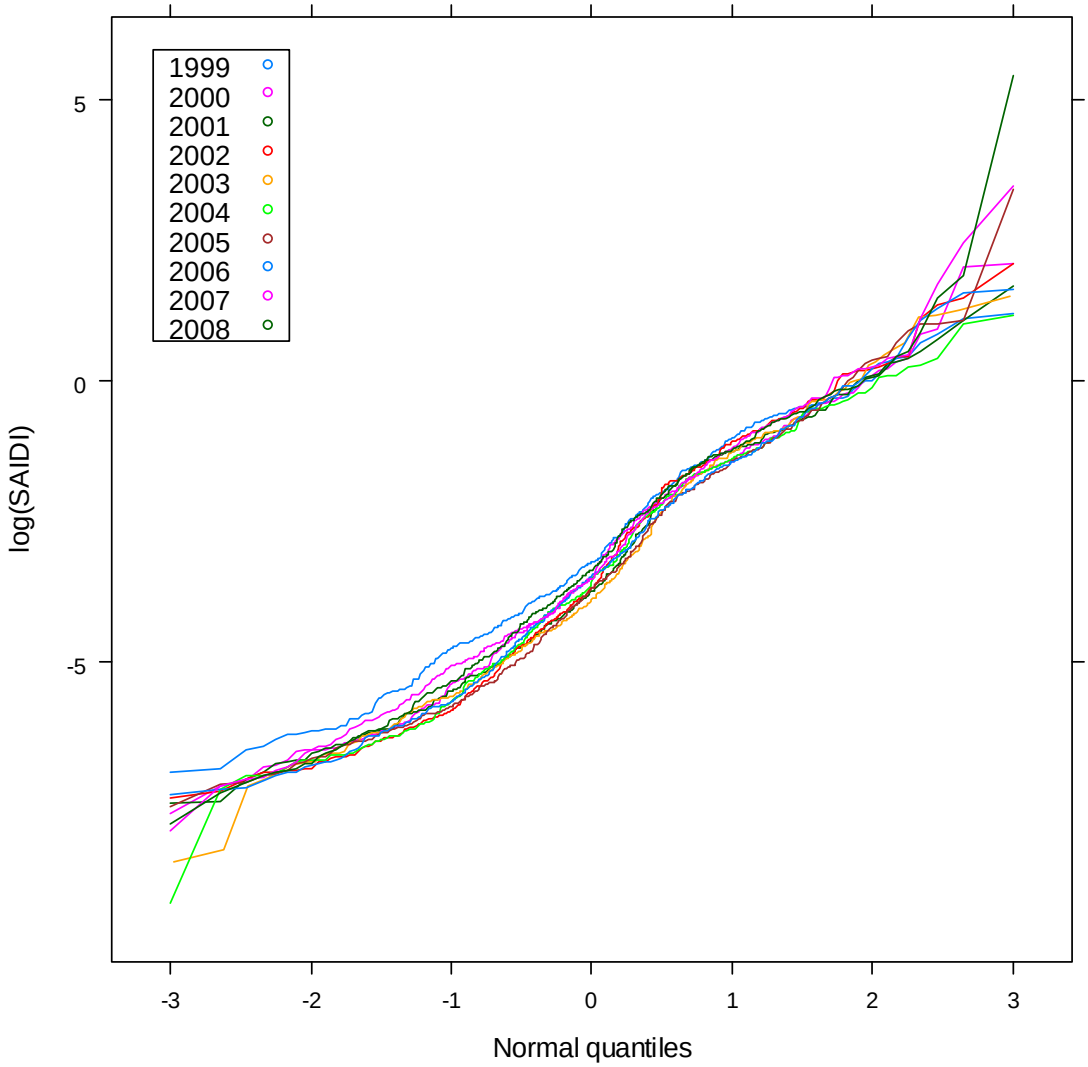

Figure 1: Normal probability plots by year, 1999-2008

However, the general shape of the curves is roughly consistent from year to year, and shows that the distribution of log(SAIDI) consistently deviates from normality.

We can also look at the data in five-year periods, since this relevant to the AER's procedure. This is shown in Figure 2. There is clearly no effective difference between the six ditributions. That is, the distribution has stayed constant over a five-year period.

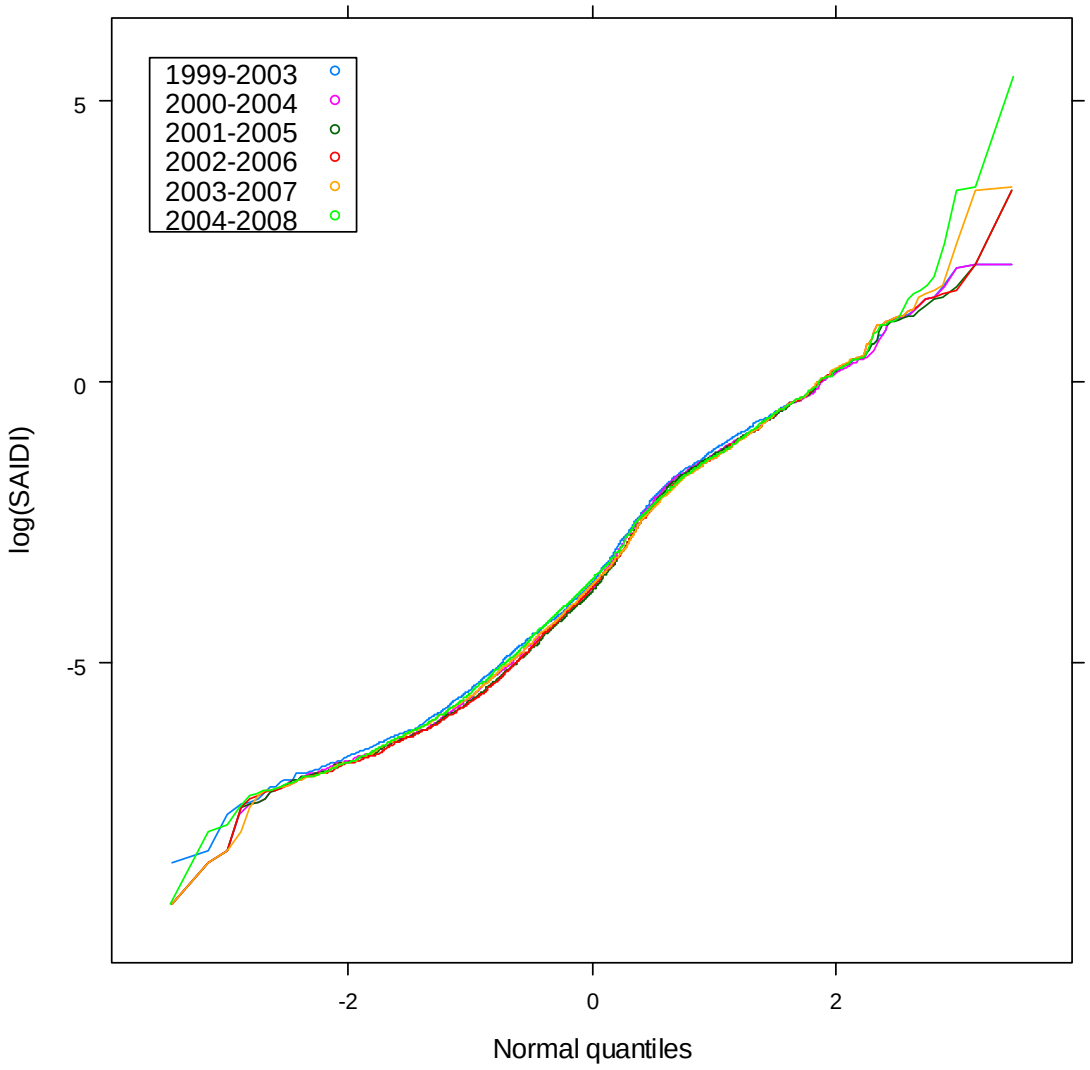

Figure 2: Normal probability plots by 5-year periods

We can carry out formal tests of normality on both the five-yearly periods for log(SAIDI) (Table 1) and the corresponding Box-Cox transformed data (Table 2).

Table 1: Significance probabilities of tests for normality of log(SAIDI)

Test	1999-03	2000-04	2001-05	2002-06	2003-07	2004-08
Anderson-Darling	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$
Cramer-von Mises	$4x10^{-7}$	$2x10^{-5}$	$8x10^{-5}$	$6x10^{-9}$	$5x10^{-10}$	$4x10^{-10}$
Lilliefors	$3x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$1x10^{-14}$	$5x10^{-11}$	$2x10^{-9}$
Shapiro-Wilk	$7x10^{-15}$	$1x10^{-15}$	$5x10^{-16}$	$1x10^{-14}$	$8x10^{-13}$	$1x10^{-12}$

Table 2: Significance probabilities of tests for normality of Box-Cox(SAIDI)

Test	1999-03	2000-04	2001-05	2002-06	2003-07	2004-08
Anderson-Darling	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$
Cramer-von Mises	$3x10^{-8}$	$8x10^{-7}$	$2x10^{-7}$	$1x10^{-9}$	$4x10^{-10}$	$4x10^{-10}$
Lilliefors	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$<2x10^{-16}$	$3x10^{-14}$	$2x10^{-12}$
Shapiro-Wilk	$2x10^{-14}$	$3x10^{-15}$	$3x10^{-15}$	$4x10^{-14}$	$7x10^{-12}$	$7x10^{-12}$
Box-Cox λ	-0.0295	-0.0294	-0.0415	-0.0375	-0.0403	-0.0359

Clearly, neither log(SAIDI) nor the Box-Cox transformed data can be regarded as normally distributed.

4. Change in SAIDI

The boxplots in Figure 3 below set out the distribution of log(SAIDI) by calendar year. The data comes from the DATA2 and DATA1 data sets.

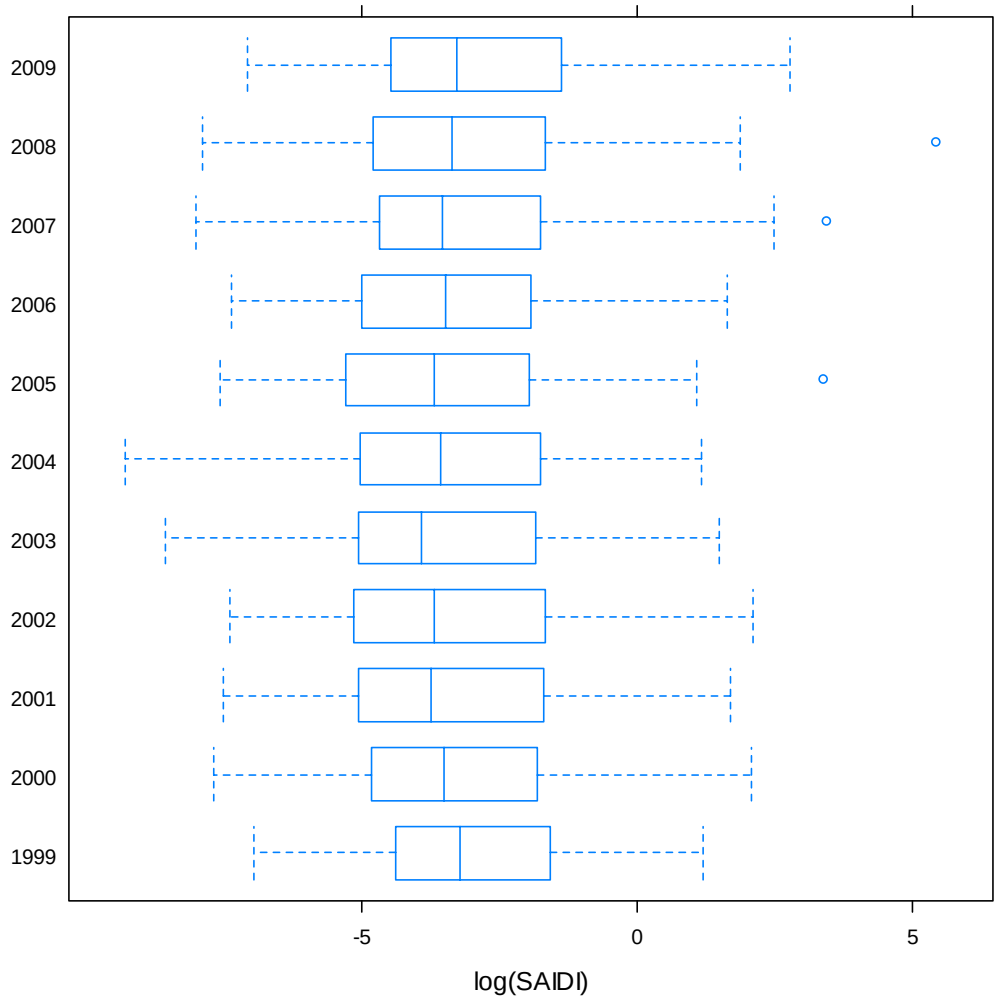

Figure 3: Annual changes in the distribution of log(SAIDI)

The median (50th quantile), 75th and 95th quantiles for SAIDI are shown in Table 3 below. Since the logarithm is a monotonic transformation, there is a direct correspondence between the quantiles of SAIDI and of log(SAIDI).

Table 3: SAIDI quantiles by year

Year	Median	75th quantile	95th quantile
1999	0.0397	0.206	0.676
2000	0.0302	0.164	0.680
2001	0.0238	0.185	0.601
2002	0.0252	0.188	0.724
2003	0.0202	0.159	0.710
2004	0.0286	0.173	0.614
2005	0.0253	0.140	0.660
2006	0.0308	0.144	0.680
2007	0.0295	0.172	0.750
2008	0.0345	0.189	0.743
2009	(0.0380)	(0.252)	(1.659)

In broad terms, the median value (the central line of the boxplots) decreased from 1999 until 2003, and then has increased since 2003. The 75th quantile (the upper limit of the 'box' of the boxplots) declined until about 2005, but has increased since then. The 95th quantile has stayed approximately steady until the current year; the marked increase for 2009 is largely due to the incomplete data for the year, and should be ignored at this stage.

So there does appear to be evidence that the distribution of SAIDI has shifted upwards since 2003 – 2005.

This is consistent with the behaviour seen in the Figure 1.

5. Rural vs. Urban feeders

This section examines the distribution of urban and rural SAIDI separately, to see if differences between the distributions are affecting the combined distribution. We use the years 2004 and 2005, since we have SAIDI data for the urban/rural split for 2005 (DATA4), and can calculate it for 2004 from DATA5 and DATA3. The reasons for ignoring other years are as follows

1999 – 2002	No consistent rural/urban split for customer numbers
2003	Three weeks SAIDI/CMOS data missing at end of year
2006 – 2008	No rural/urban split for SAIDI/CMOS

The histograms in Figure 4 show log(SAIDI) split between urban and rural areas, for both 2004 and 2005. Zero values of SAIDI have been omitted.

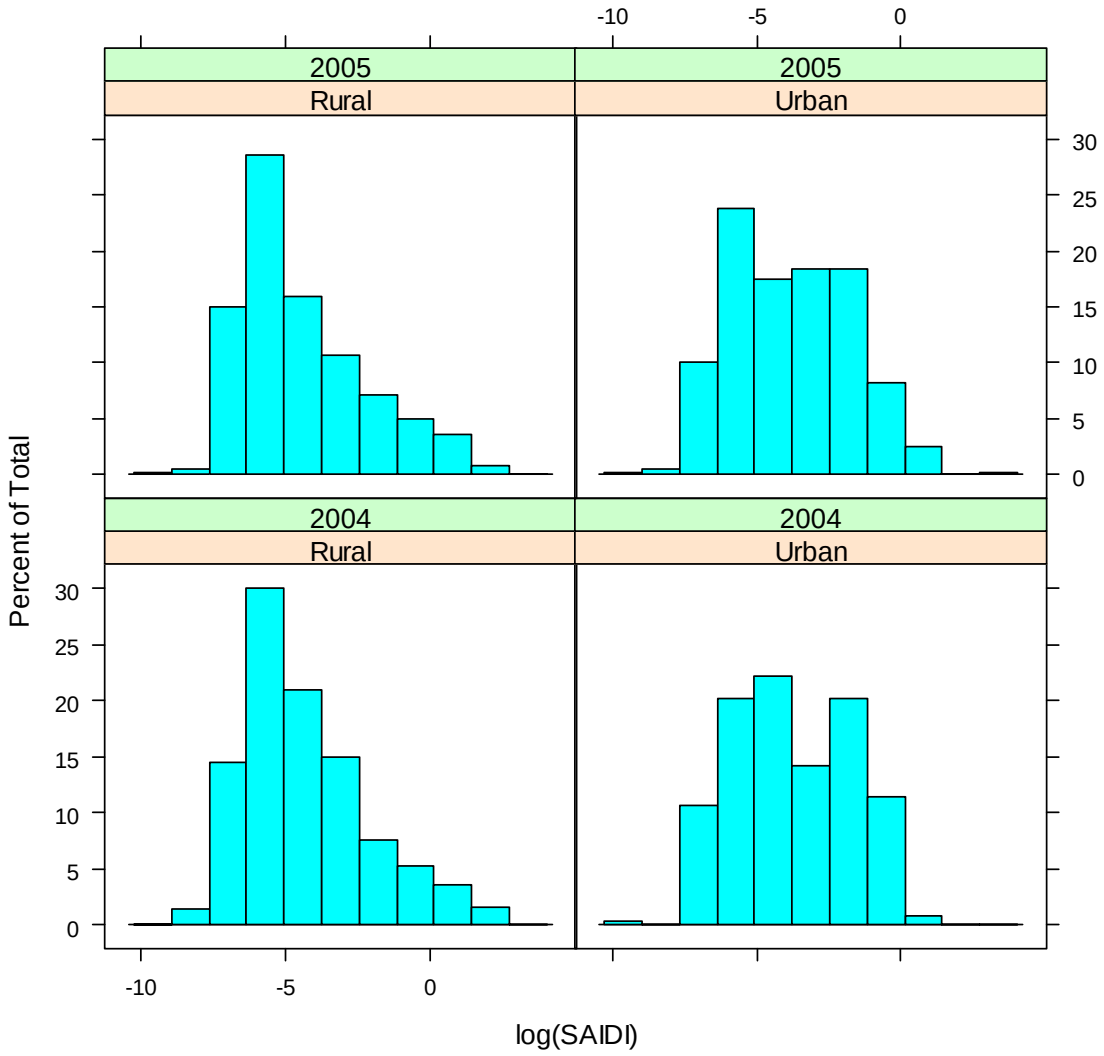

Figure 4: Histograms of log(SAIDI) for urban and rural feeders, 2004 & 2005

There are obvious differences between the rural and urban distributions: the urban distribution is more symmetric than the rural distribution, and the rural distribution has a longer upper tail. There is less difference between the two years.

The situation is emphasised by normal probability plots for the four sets of SAIDI values, shown in Figure 5. The lower pair of lines is for the two rural distributions, and the upper pair is for the urban distributions. There is little difference between the two years. While the distributions do not appear normal, there is more similarity in the lower tails than in the middle regions or the upper tail.

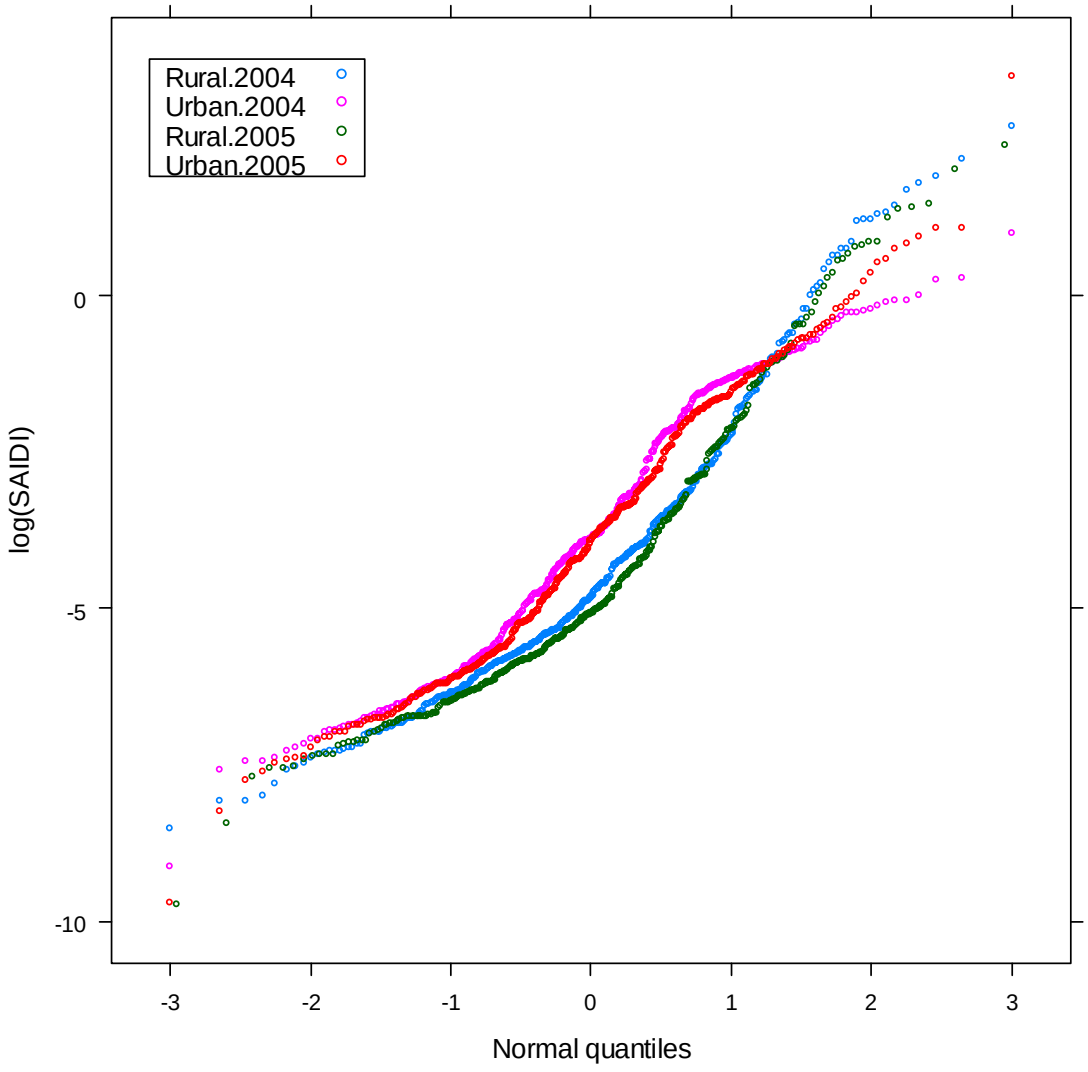

Figure 5: Normal probability plots for urban and rural feeders

We next examine the effectiveness of a Box-Cox transformation on the rural and urban components of SAIDI. The lambda values are -0.1686 for rural SAIDI and -0.0364 for urban SAIDI.

Table 4 sets out the results of normality tests for the urban SAIDI components, for both the log transform and the Box-Cox transform. Similar information for rural feeders is shown in Table 5.

Table 4: Significance probabilities for normality tests of SAIDI for urban feeders

URBAN	Log	Box-Cox
Anderson-Darling	$P = 4 \times 10^{-15}$	$P = 3 \times 10^{-14}$
Cramer-von Mises	$P = 4 \times 10^{-9}$	$P = 8 \times 10^{-9}$
Lilliefors	$P = 1 \times 10^{-9}$	$P = 2 \times 10^{-8}$
Shapiro-Wilk	$P = 1 \times 10^{-9}$	$P = 3 \times 10^{-9}$

Table 5: Significance probabilities for normality tests of SAIDI for rural feeders

RURAL	Log	Box-Cox
Anderson-Darling	$P < 2 \times 10^{-16}$	$P = 4 \times 10^{-6}$
Cramer-von Mises	$P = 4 \times 10^{-5}$	$P = 0.0002$
Lilliefors	$P < 2 \times 10^{-16}$	$P = 4 \times 10^{-5}$
Shapiro-Wilk	$P = 2 \times 10^{-16}$	$P = 6 \times 10^{-6}$

The Box-Cox transformation provides no improvement over the log transformation for the urban SAIDI values, and only marginal improvement for the rural values.

Consideration of this subdivision of the data does not appear to offer a route to normality.