Bucket Plot: A Visual Tool for Skewness and Kurtosis Comparisons

Fernanda De Bastiani\textsuperscript{a} and Dimitrios M. Stasinopoulos\textsuperscript{b} and Robert A. Rigby\textsuperscript{b} and Gillian Z. Heller\textsuperscript{c,d} and Lucas A. Silva\textsuperscript{a}

\textsuperscript{a}Statistics Department, Universidade Federal de Pernambuco  
\textsuperscript{b}STORM, London Metropolitan University, London, UK  
\textsuperscript{c}NHMRC Clinical Trials Centre, University of Sydney  
\textsuperscript{d}Department of Mathematics and Statistics, Macquarie University

Abstract. This study introduces the bucket plot, a visual tool to detect skewness and kurtosis in a continuously distributed random variable. The plot can be applied to both moment and centile skewness and kurtosis. The bucket plot is used to detect skewness and kurtosis either in a response variable, or in the residuals from a fitted model as a diagnostic tool by which to assess the adequacy of a fitted distribution to the response variable regarding skewness and kurtosis. We demonstrate the bucket plot in nine simulated skewness and kurtosis scenarios, and the usefulness of the plot is shown in a real-data situation.

1 Introduction

Traditionally statistical methodology, and especially regression methodology, was designed to detect differences in the location of the distribution of the response variable. In order to do that, the analysis concentrated on modelling the first moment (the mean) of the distribution of the response variable. The variance of the response variable, was modelled as a constant and used to evaluate whether differences in the mean were statistically significant. Other (not central parts) of the distribution were usually ignored. With the availability of larger data sets and also with the introduction of Generalized Additive Models for Location, Scale, and Shape (GAMLSS) for regression modelling, Rigby and Stasinopoulos (2005), this has changed. The practitioner can now model not only the center of the distribution but also the spread and other characteristics. Those other characteristics are often related to the shape of the distribution like the skewness and the kurtosis of the response variable. Those two higher moments can be very important, especially for studies associated with risk assessment, where the behavior in the tail of the distribution is of high interest. In this article we develop a tool for evaluating the adequacy of the modelling of the skewness and kurtosis. This tool could help practitioners to correctly assess the suitability of a theoretical distribution as far as skewness and kurtosis is concerned. Understanding the shape of the response variable distribution is a crucial and important component in data analysis.

Informally, the distribution of a single response variable is considered symmetrical if its left and right sides relative to the median resemble mirror images of one another. Kurtosis is primarily a measure of the heaviness of the tails, which is measured relative to the normal distribution. A distribution can have heavier (\textit{lepto}) or lighter (\textit{platy}) tails than the normal distribution. Existing statistical software usually has built-in functions by which to calculate the (moment) skewness and kurtosis of the response variable.

There are tools available for checking skewness and kurtosis in a response variable (or in the residuals from a fitted model). Boxplots, histograms, symmetric plots and QQ-plots are among them. Boxplots and histograms can be used as exploratory tools to assess if a data set is symmetric about the mean and possibly to detect outlier observations, but their
main purpose is to give an overview of the location, spread and shape, etc. of the underlying distribution. There are also available in the literature boxplots adjusted for skewed distributions, as proposed by Hubert and Vandervieren (2008). Symmetry plots can be used for detecting asymmetry in the response variable (or residuals). The ordered values of distances below the median are plotted against the ordered distances above the median. [More precisely, \((m - y(i))I(y(i) < m)\) is plotted against \((y(n-i+1) - m)I(y(n-i+1) > m)\) where \(m\) is the median, \(I()\) is the Heaviside function and \(y(i)\) from \(i = 1, 2, \ldots, n\) are the ordered observations.] Normal QQ-plots are useful in identifying differences in the distribution of the response variable (or residuals) compared to the normal distribution and are therefore useful in detecting a breakdown in a normal distribution assumption. The QQ-plot can also be used to check model adequacy, for instance see Augustin, Sauleaub and Wood (2012) for more details about QQ-plot in the context of generalized linear models. Note that QQ-plots are rather difficult to interpret, while the detrended QQ-plots also known as worm plots, van Buuren and Fredriks (2001) are much easier. Chapter 12 of Stasinopoulos et al. (2017) covers the use of (single and multiple) worm plots within GAMLSS.

The current study proposes a visual tool, the bucket plot, which concentrates on detecting skewness and kurtosis problems in a fitted model using any continuous distribution. This tool can be applied either at a preliminary analysis stage to suggest an appropriate distribution for the response variable, or at a post-fit stage to check the residual skewness and kurtosis from the fitted model and distribution. This tool helps to verify if assumptions of the model are being violated. For so long, transformation of the response has been used to try to transform it to a symmetric distribution which was fitted using the normal distribution, however this was often not successful. Many statistical methods have now been developed in regression analysis that can deal with different shapes of the distribution (e.g. GAMLSS). One of the challenges is to select which distribution fits better to the response variable. The development of the bucket plot has its origins in the desire to find ways to compare theoretical distributions in terms of their flexibility in modeling two different ordering measures of skewness and kurtosis—namely, the moment and centile skewness and kurtosis. New distributions appear every year in the literature; however, there is a dearth of tools by which to compare their properties to those of existing distributions. The bucket plot can be used to compare the flexibility of continuous distributions implemented within GAMLSS software in terms of their ability to model skewness and kurtosis, see Rigby et al. (2019). De Bastiani et al. (2019) can be considered the first attempt to use a bucket plot as a model diagnostic tool.

MacGillivray (1986) provides a brief historical survey of work on skewness and a wide variety of skewness measures and orderings; see also the work of MacGillivray and Balanda (1988). Balanda and MacGillivray (1990) provide several theoretical considerations of kurtosis. Horsewell and Looney (1993) discuss the diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality. Joanes and Gill (1998) compare measures of sample skewness and kurtosis and Cullen et al. (1999) present a plot for skewness and kurtosis comparison. Kim and White (2004) suggest a robust estimation of skewness and kurtosis. Brys et al. (2003) propose new measures of skewness that they claim are more robust against outlier values. Few studies in the literature discuss skewness and kurtosis, and there has been very little discussion on diagnostic methods by which to select a distribution based on skewness and kurtosis diagnostics.

The remainder of this paper is organized as follows. Section 2 discusses definitions of moment and centile skewness and kurtosis. In Section 3, six different theoretical distributions implemented in the \texttt{R} package \texttt{gamlss.dist} (Rigby et al., 2019) are compared in terms of their flexibility in modeling skewness and kurtosis. In Section 3, we introduce a bucket plot while using simulation examples that feature different combinations of skewness and kurtosis. We present an application of single and multiple bucket plots to a real dataset in Section 4. Section 5 concludes this paper.
2 Definitions for Skewness and Kurtosis

2.1 Skewness

Conceptually, the idea of skewness is straightforward, because we can think of it as a lack of symmetry; however, a rigorous definition for this term is neither obvious nor unique. The distribution of a random variable \( Y \) is defined as right-skewed (i.e., “positively skewed”) if \( Y \) is skewed more to the right than \(-Y\), according to a particular skewness ordering. The problem is that “more skew” is not uniquely defined, because there are many different theoretical skewness orderings and they are not all necessarily equivalent. MacGillivray (1986) thoroughly investigate the concept of skewness. Rigby et al. (2019) [Chapter 14] use three different skewness definitions: (i) moment skewness, (ii) centile skewness, and (iii) van Zwet skewness van Zwet (1964a,b). The current study uses moment- and centile-based skewness.

2.1.1 Moment Skewness

The moment skewness of a random variable \( Y \) is defined as:

\[
\gamma_1(Y) = \frac{\mu_3}{(\mu_2)^{3/2}},
\]  

where \( \mu_k \) is the \( k \)th central moment of \( Y \). It is also known as Pearson’s moment coefficient of skewness. Using this measure, the “moment positive skewness” is defined by \( \gamma_1(Y) > 0 \), provided that \( \gamma_1(Y) \) is finite. The distribution of \( Y_2 \) is more moment skew to the right compared to the distribution of \( Y_1 \) if \( \gamma_1(Y_2) > \gamma_1(Y_1) \), where \( \gamma_1(Y_i) \) is the moment skewness of \( Y_i \) for \( i = 1, 2 \). Unfortunately, for several heavy-tail distributions, some (or all) of their moments are either infinite or undefined; therefore, for those distributions, the moment measure of skewness may not exist. However, a centile skewness always exists and can be used as a replacement for moment skewness.

A comparison of the moment skewness for different distributions can be facilitated by monotonically transforming the moment skewness to a finite range. The transformed moment skewness with a range of values \((-1, 1)\) is defined as follows.

\[
\gamma_{1t}(Y) = \frac{\gamma_1(Y)}{(1 + |\gamma_1(Y)|)}
\]

2.1.2 Centile Skewness

The centile skewness function of a random variable \( Y \) is defined as:

\[
s_p(Y) = \frac{(y_p + y_{1-p})/2 - y_{0.5}}{(y_{1-p} - y_p)/2}
\]

for \( 0 < p < 0.5 \), where \( y_p = F_Y^{-1}(p) \) and \( F_Y^{-1}(\cdot) \) is the inverse cumulative distribution function of \( Y \). Note that centile skewness is always in the interval \((-1, 1)\), i.e. \(-1 < s_p(Y) < 1\) for all \( 0 < p < 0.5 \).

Using this measure, positive centile skewness is defined by \( s_p(Y) \geq 0 \) for all \( 0 < p < 0.5 \), with \( s_p(Y) > 0 \) for some \( p \). The distribution of \( Y_2 \) is a “more centile skew to the right” than the distribution of \( Y_1 \) if \( s_p(Y_2) \geq s_p(Y_1) \) for all \( 0 < p < 0.5 \), with \( s_p(Y_2) > s_p(Y_1) \) for some \( p \).

Note that the centile skewness function is a function of \( p \). One important case is \( p = 0.01 \), and \( s_{0.01}(Y) \) is called the tail centile skewness. The other important case is \( p = 0.25 \), which gives Galton’s measure of skewness, or central centile skewness:

\[
\gamma = s_{0.25} = \frac{(Q_1 + Q_3)/2 - m}{(Q_3 - Q_1)/2},
\]
where \( Q_1 = y_{0.25}, Q_3 = y_{0.75}, \) and \( m = y_{0.5} \) from Equation (2). Hence, \( \gamma \) is the (midquartile – median) divided by the semi-interquartile range.

The two criteria by which to compare distributions (i.e., moment and centile skewness) are not equivalent. For example, when \( Y \) belongs to a \( t \) family distribution, \([\text{that is, } Y \sim TF(\mu, \sigma, \nu), \text{ where } \mu \text{ is the location parameter, } \sigma \text{ is the scale parameter, and } \nu \text{ is the degree of freedom parameter}], \) then the moment skewness is finite only if \( \nu > 3 \), but the centile skewness always exists.

### 2.2 Kurtosis

Kurtosis characterizes the “tail” and the “shoulders” of a distribution. Informally, a distribution is considered “more kurtotic” if it has heavier tails and lighter shoulders (e.g., lighter density around its quartiles). A distribution is defined as leptokurtic (platykurtic) if it is more (less) kurtotic than the normal distribution, according to a particular kurtosis ordering. The problem is that “more kurtotic” is not uniquely defined, as there are many different kurtosis orderings that are often not equivalent. The following are two criteria by which to compare the kurtosis between two distributions.

#### 2.2.1 Moment Excess Kurtosis

The moment excess kurtosis of a random variable \( Y \) is defined by

\[
\gamma_2(Y) = \frac{\mu_4}{(\mu_2)^2} - 3
\]

where \( \mu_k \) is the \( k \)th central moment of \( Y \). The distribution of \( Y_2 \) is more moment-kurtotic than the distribution of \( Y_1 \) if \( \gamma_2(Y_2) > \gamma_2(Y_1) \), where \( \gamma_2(Y_i) \) is the moment excess kurtosis of \( Y_i \) for \( i = 1, 2 \).

Since \( \gamma_2(Y) = 0 \) when \( Y \) has a normal distribution, then when \( \gamma_2(Y) < 0 \), this indicates that \( Y \) has a moment platykurtic distribution, while \( \gamma_2(Y) > 0 \) indicates a moment leptokurtic distribution. Moreover, a monotonic transformation to scale \( \gamma_{2t}(Y) \) in the \((-1,1)\) interval is given by

\[
\gamma_{2t}(Y) = \frac{\gamma_2(Y)}{1 + |\gamma_2(Y)|},
\]

#### 2.2.2 Centile Excess Kurtosis

Another kurtosis measure is the centile kurtosis function (MacGillivray, 1986), defined as:

\[
k_p(Y) = \frac{y_{1-p} - y_p}{y_{0.75} - y_{0.25}}
\]

for \( 0 < p < 0.5 \). The distribution of \( Y_2 \) is “more centile kurtotic” than the distribution of \( Y_1 \) if

\[
k_p(Y_2) \geq k_p(Y_1), \text{ for all } 0 < p < 0.25
\]

and

\[
k_p(Y_2) \leq k_p(Y_1), \text{ for all } 0.25 < p < 0.5
\]

with \( k_p(Y_2) \neq k_p(Y_1) \) for some \( p \).

Note that condition (5) is one definition of \( Y_2 \) with heavier tails than \( Y_1 \), while condition (6) is one definition of \( Y_2 \) being more peaked than \( Y_1 \) around their medians. An important case is \( p = 0.01 \) in (4), which gives \( \delta = k_{0.01} \). The normal distribution has a centile kurtosis of \( k_{0.01} = 3.449 \). Hence, the centile excess kurtosis \( ek_{0.01} \) is given by \( ek_{0.01} = k_{0.01} - 3.449 \).

Moreover, a transformation to scale \( tk_{0.01} \) in the \((-1,1)\) interval is given by

\[
tk_{0.01} = \frac{ek_{0.01}}{1 + |ek_{0.01}|}.
\]
3 Skewness and Kurtosis Comparison using the Bucket Plot

3.1 Comparison of distributions

Here we motivate the bucket plot by a skewness and kurtosis comparison of six continuous distributions, with support on the real line $\mathbb{R} = (-\infty, \infty)$, see Chapter 16 of Rigby et al. (2019). These distributions comprise a subset of the distributions implemented in the R package `gamlss.dist` and include: (i) the exponential generalized beta type 2 (EGB2) (McDonald and Xu, 1995), (ii) the Johnson SU (JSU) (Johnson, 1949), (iii) the Skew t type 3 (ST3) (Fernandez and Steel, 1998), (iv) the skew power exponential type 3 (SEP3) (Fernandez, Osiewalski and Steel, 1995), (v) sinh-arcsinh (SHASHo) (Jones and Pewsey, 2009), and (vi) the stable distribution (SB) (Nolan, 2012). More details about these distributions can be found in Rigby et al. (2019). All six distributions have four parameters. These six distributions are also location-scale family distributions, each with the following parameters: $\mu$, the location shift parameter; $\sigma$, the scaling parameter; and $\nu$ and $\tau$, parameters primarily affecting the skewness and kurtosis of the distribution, respectively. Note that neither $\mu$ nor $\sigma$ affect the moment (or centile) skewness and kurtosis, and they are therefore held constant at values 0 and 1, respectively, in Figures 1, 2 and 3.

Figure 1 (a) shows the approximate active region of possible combinations of transformed moment kurtosis, $\gamma_2t$, and transformed (positive) moment skewness, $\gamma_1t$, for five of the aforementioned distributions. (We exclude the stable family of distribution (SB) from the moment skewness and kurtosis plots of Figure 1 (a), since moments for the stable family do not exist, in general.) Values for transformed negative moment skewness, $\gamma_1t$, constitute a mirror image of Figure 1 (a), reflected about the vertical axis at zero skewness, and while they are not plotted here, they will be used later in the bucket plot. Two-parameter location-scale distributions can be presented as points in the plot, since for fixed values of $\mu$ and $\sigma$, the moment skewness and kurtosis are fixed. The normal distribution is at point $(0, 0)$. In the plot, three-parameter location-scale distributions are presented as lines, while four-parameter location-scale distributions are presented as regions, of permissible values for skewness and kurtosis.

Since we use transformed measures for moment skewness and moment kurtosis, the regions are bounded at $(-1, 1)$ in both directions. The solid line (below all other lines) in Figure 1 (a) represents the lower bound for all possible distributions, [that is, the transformed moment skewness and kurtosis of any distribution cannot go below this line, nor above 1 in the kurtosis direction]. Figure 1 (a) is similar to the Cullen and Frey plot (Cullen and Frey, 1999), save for a fundamental difference. Our transformed plot is bounded, includes different distributions and is the same for any fixed value of the mean and variance of the five location-scale distributions. Note that negative transformed moment kurtosis ($\gamma_2t < 0$) corresponds to moment platykurtosis, while positive transformed moment kurtosis ($\gamma_2t > 0$) corresponds to moment leptokurtosis. A distribution at $\gamma_2t = 0$ is called a moment mesokurtotic distribution. Figure 1 (a) shows that the distributions JSU, ST3, and EGB2 allow only moment leptokurtosis, while SEP3 and SHASHo allow both moment platykurtosis and moment leptokurtosis. For very low moment kurtosis, SEP3 is the only distribution that allows a range of moment skewness, while for higher moment kurtosis, SHASHo allows the largest range of moment skewness.

Figure 1 (b) shows for all six distributions the regions of transformed excess centile kurtosis, $tk_{0.01}$, and (positive) centile skewness, $s_{0.25}$. Again, the flexibility of each distribution in modeling centile kurtosis and centile skewness is seen in the area lying above each distribution curve. The curves are the (approximate) boundaries; Rigby et al. (2019) [Chapter 16] provide an example of why the curves may be approximate. Again, the corresponding regions for negative skewness are given by the reflections of Figure 1 (b) about the vertical
axis at zero skewness. The normal distribution is at point \((0, 0)\) in Figure 1 (b). Transformed centile excess kurtosis below zero is centile platykurtic, while above zero is centile leptokurtic. Hence SB, EGB2, and JSU have only centile leptokurtic distributions. The \(\text{SEP3}\) is the only distribution that allows for very low centile kurtosis, while for higher centile kurtosis \(\text{SHASH0}\) allows the largest range of centile skewness, and it looks to be the most versatile distribution in terms of the combinations of centile skewness and kurtosis. The \(\text{SB, JSU, and ST3}\) distributions for a given transformed centile kurtosis often have a relatively restrictive centile skewness. The stable distribution, \(\text{SB}\), has the most restricted region of centile skewness for a given transformed centile kurtosis. For all distributions—except for \(\text{EGB2}\)—the range of possible centile skewness increases with the transformed centile kurtosis.

### 3.2 The bucket plot for diagnostic comparison of fitted models

The bucket plot is a graphical tool to help to select a distribution that best fits the response variable in terms of its distribution shape (specifically skewness and kurtosis). It tells us which of the models should not be used because they do not adequately fit the response distribution shape. The bucket plot is based on the plots shown in Figures 1 (a) and (b), but includes both negative and positive skewness. For a given response variable, the bucket plot is useful for checking the adequacy of the fitted distribution in terms of skewness and kurtosis. If the aim is to fit a marginal distribution to the response variable, the bucket plot can be used as an exploratory tool to select the correct marginal distribution. In a fitted regression analysis, the bucket plot can be used as a diagnostic tool, using residuals to check the adequacy of the distribution (and the explanatory variable models for its parameters). In this case, normalized quantile residuals are used (Dunn and Smyth, 1996). If the model is correct, then the true values of these residuals have a standard normal distribution (even when the model distribution is not normal), as discussed in Chapter 12 of Stasinopoulos et al. (2017).

In a bucket plot, the moment or centile skewness and kurtosis of the response variable (or residuals from a fitted model for the response variable) are plotted as a point in the moment or centile skewness–kurtosis plot in Figures 1 (a) and (b), respectively. According to the position on which the point falls, judgment can be made about a suitable distribution for the response variable (or whether skewness and kurtosis are adequately modeled). However, bucket plots can only detect skewness and/or kurtosis: they are not designed to check whether location and scale parameters are adequately modeled. For this purpose, we can use the worm plots of van Buuren and Fredriks (2001) or the \(Q\) and \(Z\) statistics of Royston and Wright (2000). The \texttt{gamlss} packages of Stasinopoulos et al. (2017) provide tools, for model diagnostics in a distributional regression setup, and most of the functions used to produce the plots presented in this paper are implemented in packages \texttt{gamlss} and \texttt{gamlss.ggplots}.

In order to demonstrate the bucket plot, we simulated 1000 random observations from the skew exponential power type 3, \(\text{SEP3}\) \((\mu, \sigma, \nu, \tau)\), distribution with fixed \(\mu = 0\) and \(\sigma = 1\) (since \(\text{SEP3}\) is a location-scale distribution and so the skewness and kurtosis measures are independent of the location and scale), but with different values of \(\nu\) and \(\tau\).

Table 1 shows the different values of the skewness parameter \(\nu\) and the kurtosis parameter \(\tau\) used in the simulation. The table also shows the sample transformed moment skewness, \(\gamma_{1t}(Y_i)\), the sample transformed moment excess kurtosis, \(\gamma_{2t}(Y_i)\), the sample tail centile skewness \(s_{0.01}(Y_i)\), and the sample transformed centile excess kurtosis \(t_{k0.01}(Y_i)\), for \(i = 1, \ldots, 9\). The distribution used for the simulation is \(\text{SEP3}\) [since, according to Figures 1 (a) and (b)], \(\text{SEP3}\) provides a flexible combination of skewness and kurtosis. Note that the distribution of \(Y_5\) is a special case since \(\text{SEP3}\) \((0, 1, 1, 2)\) is the standard normal distribution.
**Figure 1** The regions of combinations of (a) transformed moment excess kurtosis, $\gamma_2^t$, and (positive) transformed moment skewness, $\gamma_1^t$, for five distributions: exponential generalized beta type 2 (EGB2), Johnson’s Su (JSU), skew t distribution type 3 (ST3), skew exponential power type 3 (SEP3), and Sinh-Arcsinh (SHASHo), and for all distributions (black line boundary). (b) transformed centile excess kurtosis and (positive) tail centile skewness, for six distributions stable SB, EGB2, JSU, ST3, SEP3, and SHASHo, and for all distributions (black line rectangle boundary).
for each randomly generated sample. Samples distribution. Figure 2 shows the simulating probability density functions as well as the histogram (first row), mesokurtotic (second row), and leptokurtotic (third row) from the (column), symmetric (second column), and right skewness (third column) and platykurtotic Table 1 The distribution parameters values for ν and τ and the sample transformed moment skewness (γ1t(Yi)), transformed moment excess kurtosis (γ2t(Yi)), tail centile skewness (s0.01(Yi)) and transformed centile excess kurtosis (tk0.01(Yi)), for i = 1, . . . , 9, from simulated samples of size n = 1000 generated from the four parameter skew power exponential type 3 (SEP3(µ, σ, ν, τ)) distribution with fixed µ = 0 and σ = 1.

<table>
<thead>
<tr>
<th>Sample Yi</th>
<th>ν</th>
<th>τ</th>
<th>γ1t(Yi)</th>
<th>γ2t(Yi)</th>
<th>s0.01(Yi)</th>
<th>tk0.01(Yi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>0.5</td>
<td>4</td>
<td>−0.245</td>
<td>−0.388</td>
<td>−0.194</td>
<td>−0.469</td>
</tr>
<tr>
<td>Y2</td>
<td>1</td>
<td>4</td>
<td>−0.012</td>
<td>−0.476</td>
<td>0.005</td>
<td>−0.520</td>
</tr>
<tr>
<td>Y3</td>
<td>2</td>
<td>4</td>
<td>0.242</td>
<td>−0.437</td>
<td>0.212</td>
<td>−0.513</td>
</tr>
<tr>
<td>Y4</td>
<td>0.5</td>
<td>2</td>
<td>−0.399</td>
<td>0.102</td>
<td>−0.309</td>
<td>−0.344</td>
</tr>
<tr>
<td>Y5</td>
<td>1</td>
<td>2</td>
<td>0.038</td>
<td>0.0785</td>
<td>0.042</td>
<td>0.086</td>
</tr>
<tr>
<td>Y6</td>
<td>2</td>
<td>2</td>
<td>0.420</td>
<td>0.130</td>
<td>0.363</td>
<td>−0.199</td>
</tr>
<tr>
<td>Y7</td>
<td>0.5</td>
<td>1.5</td>
<td>−0.563</td>
<td>0.745</td>
<td>−0.444</td>
<td>0.402</td>
</tr>
<tr>
<td>Y8</td>
<td>1</td>
<td>1.5</td>
<td>0.041</td>
<td>0.452</td>
<td>−0.007</td>
<td>0.178</td>
</tr>
<tr>
<td>Y9</td>
<td>2</td>
<td>1.5</td>
<td>0.509</td>
<td>0.589</td>
<td>0.391</td>
<td>0.213</td>
</tr>
</tbody>
</table>

We see in Figure 2 nine simulated samples having all combinations of left skewness (first column), symmetric (second column), and right skewness (third column) and platykurtotic (first row), mesokurtotic (second row), and leptokurtotic (third row) from the (SEP3) distribution. Figure 2 shows the simulating probability density functions as well as the histogram for each randomly generated sample. Samples Y1, Y2, and Y7 are platykurtotic (thin tails); Y4, Y5, and Y6 are approximately mesokurtotic; and Y7, Y8, and Y9 are leptokurtotic (heavy tails). Samples Y1, Y4, and Y7 have negative skewness; Y2, Y5, and Y8 are symmetric; and Y3, Y6, and Y9 have positive skewness.

Figure 3 (a) shows the moment bucket plot for the single simulated sample Y2. A moment bucket plot consists of the following components.

1. A bucket shape region (the black continuous line) which is the permissible transformed moment excess kurtosis and transformed moment skewness region of all possible distributions.
2. A single point marked as Y2, in the lower part of the bucket, indicating the observed transformed moment excess kurtosis γ2t and transformed moment skewness γ1t of sample Y2. The position of the sample within the bucket indicates whether there is leptokurtic (upper half) or platykurtic (lower half) moment excess kurtosis, and negative (left half) or positive (right half) moment skewness. Note that the point (0, 0) for (γ1t, γ2t) represents the values of the normal distribution. Sample values falling in the middle close to (0, 0), i.e. zero moment skewness and zero moment excess kurtosis, are compatible with the normal distribution.
3. A shaded elliptic region around the point (0, 0) indicated a 95% region for γ1t(Yi) and γ2t(Yi), based on the Jarque–Bera test (Jarque and Bera, 1980). If a sample point falls in the shaded region, the null hypothesis of the Jarque–Bera test that γ1t(Yi) = 0 and γ2t(Yi) = 0 simultaneously is not rejected; otherwise, it is rejected. Therefore, the bucket plot provides a visual means of checking Jarque–Bera test results.
4. A cloud of points around the point Y2, the transformed skewness and kurtosis of the sample. The cloud points are transformed skewness and excess kurtosis measures from 99 nonparametric bootstrap samples from variable Y2. Each bootstrap sample had the same size as the original Y2 variable. The cloud provides an impression of the variability associated with estimating the transformed moment skewness and moment excess kurtosis.

Values for the transformed skewness and excess kurtosis, taken from a single sample, that fall outside the Jarque–Bera test confidence region indicate that the skewness and kurtosis in
the variable is significantly different from that of a normal distribution, while values inside indicate that that the normal distribution might be appropriate. Similarly, if instead of a sample, we use the (normalized quantile) residuals from a fitted model for a response variable, and if the transformed skewness and excess kurtosis of the residuals fall outside the Jarque–Bera test confidence region, then the skewness and kurtosis in the response variable has not been accounted for, and some kind of correction to the model is needed. This usually involves fitting a distribution for the response variable which allows modelling of the skewness and kurtosis.

For the simulated sample sample $Y_2$, the transformed skewness and excess kurtosis, was below the Jarque–Bera test confidence region, indicating evidence for platy-kurtosis in the sample. Since in terms of skewness the point $Y_2$ falls close to zero, there is no significant skewness in the sample.
Figure 3 (b) shows the corresponding centile tail bucket plot for the same simulated sample $Y_2$, similar to moment plot in Figure 3 (a). Figure 3 (b) plots the transformed centile excess kurtosis $tk_{0.01}$ against the tail centile skewness $s_{0.01}$. Note that the range of all permissible points is a black continuous rectangle line. The shaded elliptic region around the point $(0,0)$ indicates a 95% region for $(tk_{0.01}, s_{0.01})$. Since there is no equivalent Jarque–Bera test, instead 1000 samples of length $n$ were generated from a normal distribution and their values for $(k_{0.01}, s_{0.01})$ calculated, from which an ellipse containing 95% of the points (with the 95% lowest values of Mahalanobis’s distance from the mean) was obtained, and then the ellipse was transformed from $k_{0.01}$ to $tk_{0.01}$ giving a 95% region for $(tk_{0.01}, s_{0.01})$.

Note that the conclusion about the skewness and kurtosis for sample $Y_2$ in the centile bucket plot of Figure 3(b) is similar to the one obtained from the moment bucket plot of Figure 3(a). There is significant platykurtosis in the sample $Y_2$, but no significant skewness. Figures (e) and (f) show the moment and centile bucket plots, respectively, using the residuals from the 9 respective fitted constant models.

Figures 3(c) and (d) shows all 9 samples simultaneously within a single bucket plot. Panel 3(c) shows the moment bucket, while panel 3(d) shows the centile bucket plot. Each point $Y_i$, for $i = 1, \ldots, 9$ plotted in Figures 3 (c) and (d), falls in its expected region; this correctly indicates whether skewness and/or kurtosis are present in the sample. For example, sample $Y_3$ with right-skew platykurtic characteristics falls in the lower-right quadrant, while sample $Y_7$ with left-skew leptokurtic characteristics falls in the upper-left quadrant. The only sample that falls in the Jarque–Bera test confidence region, as expected, is the sample $Y_5$. Given the bucket plot for a specific sample, an appropriate distribution with the correct type of skewness and/or kurtosis can be chosen and fitted to the data.

Figure 3 (e) and (f) show the moment and centile bucket plot of residuals, respectively, from nine fitted models, using the true model SEP3 distribution and constant models fitted for all the distribution parameters, to the nine simulated samples. For the moment bucket plot, the fitted model residuals skewness and kurtosis points for all nine samples are within the Jarque–Bera test 95% region as expected. For the centile bucket plot, the fitted model residuals skewness and kurtosis points for all nine samples also fall within the 95% region, as expected.

4 Application to London Bike-Sharing Data

Bike sharing programs are becoming increasingly popular as a new form of transport, and have been claimed to bring some benefits. Zhang and Mi (2018) presented some of the environmental benefits of bike sharing.

These data relate to the Santander bike-sharing scheme in London, a public bicycle hire scheme. The data are from January 1, 2015 to December 31, 2016, when the sponsorship of the scheme was transferred from Barclays to Santander. It comprises 17414 distinct hourly observations (Mavrodiev, 2019). The variables are cnt, the number of bike hires (response variable), $t_1$: the real temperature (Celsius), $t_2$: “feels like” temperature (Celsius), hum: percent humidity, wind_speed: wind speed (km/hour), weather_code: a factor (with levels, $1$ = clear, $2$ = scattered cloud/few clouds, $3$ = broken clouds, $4$ = cloudy, $7$ = rain/light rain shower/light rain, $10$ = rain with thunderstorm, $26$ = snowfall, and $94$ = freezing fog), is_holiday: a binary factor as to whether a holiday or not, is_weekend: a binary factor as to whether a weekend, season: a factor for seasons ($0$= spring, $1$=summer, $2$=fall, and $3$=winter), and time_of_day: hour of the day (a factor with level 0 to 23). It is available at https://www.kaggle.com/hmavrodiev/london-bike-sharing-dataset.

Figure 4 shows the distribution of the response variable cnt, where the positive skewness and right tail heaviness are clear. The response is a count variable, but because of its wide
Figure 3  Moment and centile bucket plots for the nine simulated samples, listed in Table 1 for the four parameter skew exponential power type 3 (SEP3 $(\mu, \sigma, \nu, \tau)$) distribution with fixed $\mu = 0$ and $\sigma = 1$, and $\nu$ and $\tau$ values given in Table 1. (a) moment and (b) centile bucket plot for a single sample $Y_2$; (c) moment and (d) centile bucket plot for all 9 samples $Y_1, Y_2, \ldots, Y_9$, simultaneously; (e) moment and (f) centile bucket plot using the residuals from the 9 respective fitted SEP3 models (with constant parameters). The central shaded region in each bucket plot is a 95% acceptance region for jointly testing zero skewness and zero excess kurtosis (assuming a normal random sample with the same sample size as each $Y_i$, i.e. $n=1000$ here). The curves in the bucket plots (a), (c) and (e) correspond to the curves in Figure 1(a) reflected about the vertical axis at the origin. Similarly the curves in (b), (d) and (f) correspond to the curves in Figure 1(b). The labels $Y_i$ for $i = 1, \ldots, 9$ in Figure 3(a), (c) and (e) are centred at their respective sample transformed moment skewness and transformed moment excess kurtosis, and in Figure 3(b), (d) and (f) at their respective sample tail centile skewness and transformed centile excess kurtosis. The coloured scatter cloud of points around each panel $Y_i$ are at the corresponding values of skewness and kurtosis from 99 nonparametric bootstrap samples obtained from the original $Y_i$ sample, for $i = 1, \ldots, 9$. 
range of values (0 to 7860), it is treated in the current analysis as a nonnegative continuous variable. There is a single observation having \( \text{cnt}=0 \); we excluded this observation from the data during analysis, as including it would necessitate the use of a “mixed” response zero-adjusted distribution. Additionally, we did not consider the explanatory variable “feels like” temperature \( t_2 \) in the analysis, given its high correlation with real temperature \( t_1 \).

There is a good reason why GAMLSS should be used here instead of the more conventional tools like linear models, GLM’s or GAM’s. With 17414 observations for the response it is unlikely that one or two parameter distributions used in traditional models will fit adequately. Also, as we mention earlier, traditional methods concentrate on the ‘centre’ of the distribution (in fact the mean). Theoretically, while the exponential family assumption of GLM and GAM provides asymptotically consistent estimates for the mean, inference in selecting explanatory variables for the mean may be unreliable (and even if the correct variables are selected for the mean, their parameter estimators will be inefficient) if the variance is modeled incorrectly. Also it is unlikely, with a response variable like this one, that the interest would lie only in shifts in the location (centre) of the distribution. Conditions where the demand for bikes is rather unusual are of interest. That is, low and extreme demand and the probabilities of such events to occur are important. GAMLSS models, given that an adequate distribution is fitted, provide information about the behavior of all characteristics of the distribution of the response variable. Concentrating on a specific quantile of the distribution using a quantile regression approach, only provides a partial picture of the distribution. GAMLSS, in general, needs more assumptions than for example quantile regression, but diagnostic tools, like the bucket plots based on the residuals, exist for checking those assumptions. In general, notice, that the more assumptions you have the more tools exist for checking those assumptions. This is, because the assumptions put more features in the model and that makes it easier to check departures from those features.

![Histogram of the response variable cnt, the number of bike hires, from London bike-sharing data.](image)

Figure 4 *Histogram of the response variable \( \text{cnt} \), the number of bike hires, from London bike-sharing data.*
We start by fitting two-, three-, and finally four-parameter distributions, all defined on the positive real line that are very flexible in terms of the shape of the distribution and can deal with a variety of characteristics of the data. The two-parameter distributions are the gamma (GA), inverse Gaussian (IG), inverse gamma (IGAMMA), Weibull (WEI), and log-normal (LOGNO). The three-parameter distributions are the generalized gamma (GG) (Lopatatzidis and Green, 2000), generalized inverse Gaussian (GIG) (Jørgensen, 1982), and the Box–Cox Cole and Green distribution (BCCGo) (Cole and Green, 1992). The four-parameter distributions are the Box–Cox power exponential (BCPEo) (Rigby and Stasinopoulos, 2004), Box–Cox t (BCTo) (Rigby and Stasinopoulos, 2006), generalized beta type 2 (GB2) (McDonald, 1984), log-SHASHo (LSHASHo), log-ST3 (LST3), log-Johnson SU (LJSU), and log-SEP3 (logSEP3). The last four distributions were exponentially transformed versions of the equivalent distributions defined on the real line and are therefore defined on the positive real line. (The “log” name was used to conform with the name of the log normal distribution.) These “log” distributions were created using the gamlss function gen.Family.

Figure 5 shows the fitted scaled Akaike information criterion (sAIC) for different fitted distributions. The scaling is done by setting the scaled AIC of the worst fitting distribution model (IGAMMA, AIC=240579, sAIC=0) to zero and the best (GB2, AIC=221020, sAIC=1) to one. The scaling of the rest is done according to how well the AIC performed compared with the two extreme models.

According to the scaled AIC given in Figure 5, the three best fits are models with the distributions GB2 (AIC=221020), LJSU (AIC=226554) and BCTo (AIC=226903). The bucket plots shown in next section help to check the adequacy of the fitted models in terms of skewness and kurtosis.

4.1 A Single Bucket Plot

Figure 6(a) and (b) show, respectively, the moment and centile bucket plots described in Section 3.2 applied to the residuals from the fitted models. In fact none of the fitted distributions showed an adequate fit for the residual moment/centile skewness and kurtosis. Let us focus on the top three performing distributions according to AIC: GB2, LJSU and BCTo. In the moment bucket plot, the BCTo model was closest to the (0,0) benchmark point in Figure 6(a), but ranked a poor third in the order of scaled AIC (Figure 5). Meanwhile GB2 and LJSU failed badly in moment kurtosis modelling, but were ranked first and a poor second, respectively, in the order of scaled AIC (Figure 5). We will attempt to explain this discrepancy between AIC and residuals later.

The centile bucket plot in Figure 6(b) shows a different picture from the moment bucket plot of Figure 6(a). In terms of centile kurtosis the distributions GB2 and LJSU have moved to a less extreme leptokurtic position. The explanation is that moment kurtosis of the residuals is heavily affect by very extreme residuals, while the centile kurtosis is not since it focusses on the 1% and 99% centiles, but not beyond those centiles.

The difference between the plots in Figure 6 (a) and (b) is not surprising, since the measures for skewness and kurtosis for moments, in equations (1) and (3), and for centiles in equations (2) and (4) are different. Note that a single unusual observation can greatly effect the moment (skewness and kurtosis) measures, but not the centile measures since 1% of cases in each tail are left out from the calculations. It appears that only a minority of extreme
residuals account for the extremes in skewness and especially kurtosis in the moment bucket plot.

Let us consider the difference in performance between the AIC (used in Figure 5) and the residuals which are used for the bucket plot (in Figure 7). A bucket plot for a fitted model plots the moment (or centile) skewness and kurtosis of the normalized quantile residuals (or z-scores) of the models. If the distribution (and its parameter models) fit the response variable adequately, we expect the z-score to behave as a normally distributed variable, and therefore its skewness and excess kurtosis should be close to zero. Note that the actual position of any model in the bucket plot is determined not only on how well the distribution models the skewness and kurtosis of the response variable, but also on whether the parameters affecting the skewness and kurtosis of the distribution have been modeled appropriately using explanatory variables.

The AIC depends on the global deviance $D = \sum_{i=1}^{n} d_i = -2 \sum_{i=1}^{n} \log f(y_i)$ where $f(y_i)$ for $i = 1, 2, \ldots, n$ is the fitted probability density function (pdf) for observation $i$ and $d_i$ is the corresponding deviance increment for observation $i$. However, the bucket plot depends on the normalized quantile residuals (or z-scores) for the fitted model, which are defined as $\Phi^{-1}(F(y_i))$ where $\Phi^{-1}$ is the inverse cumulative distribution function (cdf), (or quantile function), of the standardised normal distribution and $F(y_i)$ for $i = 1, 2, \ldots, n$ is the fitted cumulative distribution function (cdf) of the response. Therefore the AIC depends on the fitted pdf, while the residuals (and consequently the bucket plot) depend on the fitted cdf. An observation with a small value of the fitted pdf $f(y_i)$ will have a large deviance increment $-2\log(f(y_i))$ and therefore be influential on the AIC. However, the same observation may not have a large z-score because its fitted cdf $F(y_i)$ may not be so extreme, especially if beyond $y_i$ the fitted distribution tail is heavy.

In the above analysis of the bike counts, the GB2 model gave by far the best AIC, while the BCTo performed best in the single moment bucket plot. The explanation appears to be that the GB2 model fits the response variable bike count best in the middle part of its distribution, while BCTo fits best in the tails of its distribution. This is highlighted by Figure 7(a).
Figure 6 Transformed (a) moment and (b) centile tail bucket plots for some of the fitted models; WEI: Weibull, LST3: log skew t, LSHASH: log sinh-arcsinh, LSEP3: skew power exponential type 3, LOGNO: log normal, LJSU: log Johnson’s Su, IGAMMA: inverse gamma, IG: inverse Gaussian, GIG: generalized inverse Gaussian, GG: generalized gamma, GB2: generalized Beta type 2, BCTo: Box-Cox t, LBCPE: log Box-Cox power exponential, BCCG: Box-Cox Cole and Green.

Figure 7 (a) QQ-plot of the residuals from the GB2, generalized Beta type 2, and BCT, Box-Cox t, models (b) diff - deviance increment differences (between the models GB2 and BCTo) plotted against the cnt, the number of bike hires.

which gives a QQ-plot of the residuals from the GB2 and BCTo models, and also in Figure 7(b) where the difference in deviance increment $d_i$ between the fitted GB2 and BCTo models is plotted against the bike count $cnt_i$ for $i = 1, \ldots, n$. Clearly from Figure 7(b) the BCTo model fits much better for large values of the bike count, since their difference in deviance increments are large and positive for large bike counts. however generally the GB2 model fits better since 62% of the differences are negative, and the average difference is -0.382. In conclusion the GB2 model fits much better overall according to AIC, but the BCTo model fits better in the tails and hence performed better in the single moment bucket plot, especially for the moment kurtosis of its residuals. [Further investigation shows that there are some days
with exceptionally high bike counts, which we believe may be due to London Transport underground tube strikes, so a further analysis could include an extra explanatory binary factor indicating strike days. Further analysis could also investigate whether the bike count depends on interactions among the explanatory variables.

In general, a balance must be struck regarding what is most important in fitting each specific dataset. With the London bike-sharing data, no single fitted model (at least from the ones considered here) adequately fits the skewness and kurtosis of the response variable. One must therefore consider what is important in the analysis: the single bucket plot explains only one aspect of the information coming from the data and the fitted model—namely, the residual skewness and kurtosis—and it should be used as a diagnostic to indicate that a model is not adequate when the resulting bucket plot is found to show inadequate residual skewness and/or kurtosis.

Another very important aspect of the bucket plot, which it shares with other diagnostic tools like the worm plot of van Buuren and Fredriks (2001), is the fact that while in a single bucket plot (i.e. using all the residuals) the distribution may look adequate (as far as skewness and kurtosis are concerned), it could be the case that this is not true for specific regions of the explanatory variable space. Multiple-bucket plots, as described in the next section, should also be investigated to determine whether the dependence of the response variable distribution on one or more explanatory variables is inadequately modelled.

4.2 Multiple-Bucket Plots

Thus far, we have used the bucket plot to highlight skewness and kurtosis globally in the residuals. It is also important to use the bucket plot to investigate specific regions of the explanatory variable space and to identify whether those regions require better modeling for skewness and kurtosis. In particular, a specific region of the explanatory variable space may be a specific value of an explanatory factor, a specific range of a quantitative explanatory variable, or a specific combination of values or ranges of two or more explanatory variables. For example, in time-series data, the process of generating the response variable occasionally changes over time from one without skewness and/or excess kurtosis to one with. Multiple-bucket plots split over time would be able to check for such a change and highlight whether skewness and (positive or negative) excess kurtosis are present, and for which time period. For spatial data, skewness and/or (positive or negative) excess kurtosis may be present in some areas of the two-dimensional space but not in others; two-dimensional multiple bucket plots can be useful in identifying the areas.

Here, we demonstrate multiple bucket plots for the London bike-sharing data, where we can split the bucket for different temperatures, seasons, or any other explanatory variables of interest. Figure 8 shows a multiple-bucket plot that compares the fitted BCTo and GB2 models (described in the previous section) in four (non-overlapping) ranges of the temperature covariate $t_1$; the models are denoted as B and G, respectively. The BCTo model fits better in terms of skewness and kurtosis for all four ranges of temperature values.

Figure 9 is a two-dimensional multiple-bucket plot for the fitted models BCTo and GB2 according to the factors season and temp (the latter of which is the continuous variable $t_1$ cut into the intervals $[-1.5, 8)$, $(8, 12.5]$, $(12.5, 16]$ and $(16, 34]$). The number of observations for each cell is shown in the bottom–left corner of each individual bucket plot. One cell in Figure 9 contains zero observations and is empty, while the top right cell has only 2 observations and should be ignored. In general, the BCTo distribution performs better than the GB2 with respect to the residual moment skewness and excess moment kurtosis. However, the residual moment skewness and excess moment kurtosis of the BCTo distribution are significantly different from the $(0,0)$ values in some of the buckets—for example, when
the season is 1 (summer) and the temperature is in the interval \((12.5, 16]\). This indicates that the skewness and kurtosis of the response variable (i.e., the number of bike hires) were not modelled adequately by the \(BCTo\) model on cooler hours of summer days or nights.

![Multiple-bucket plot](image.png)

**Figure 8** Multiple-bucket plot for models generalized beta type 2 - \(GB2\) (denoted by \(G\)) and Box Cox t - \(BCTo\) (denoted by \(B\)), according to the explanatory variable temperature \(t1\) cut into the intervals \([-1.5, 8], (8, 12.5], (12.5, 16]\) and \((16, 34]\).

### 5 Conclusions

We propose a visual tool, namely, the (moment or centile) bucket plot, to investigate skewness and kurtosis. The moment bucket plot is based on a theoretical plot of the regions of possible transformed excess moment kurtosis against the transformed moment skewness for important four-parameter continuous distributions, as described in Section 3. A primitive version of the bucket plot was added to the package `gamlss` packages after the publication of Rigby et al. (2019). There are currently two different ‘full’ versions of the bucket plots available from the authors. The first is using the standard R graphical facilities while the second uses the package `ggplot2`, Wickham (2011). Both versions should be available within packages `gamlss.dist` and `gamlss.ggplots`. The bucket plot can be applied to any regression model.
where a theoretical response distribution (continuous or discrete) is assumed and where the normalized quantile residuals are obtained. Options are provided to be able to plot either the centile (central or tail) skewness against the centile kurtosis, or the moment skewness against the moment kurtosis. (In this paper, we concentrated on the centile tail skewness, rather than centile central skewness.) The centile skewness and kurtosis can be used for distributions that do not have moments.

The bucket plot can be used at a preliminary analysis stage to determine an appropriate distribution for the response variable, or after a distribution model has been fitted as a diagnostic tool to check whether skewness and kurtosis are inadequately fitted in the model. The moment bucket plot is a tool for residual analysis, and is an intuitive and graphical way to interpret the Jarque-Bera test. One can use multiple-bucket plots, during the model diagnostic checking stage, to detect those parts of the data, (i.e. those regions of the explanatory variable space), where the skewness and excess kurtosis of the response variable are not properly modelled. The use of multiple-bucket plots could also be beneficial in time-series or spatial data analysis, where the adequacy of the modelling of the skewness and kurtosis over time or space can be checked.

**Figure 9** Multiple-bucket plot for models generalized beta type 2 - GB2 (denoted by G) and Box Cox t - BCTo (denoted by B), according to the factors season (0=spring, 1=summer, 2=fall, 3=winter) and temp or t1 for temperature cut into the intervals [−1.5, 8], (8, 12.5], (12.5, 16] and (16, 34].
We also conclude that the bucket plot and AIC can give different information about the fitted model, and both are important. We note that the methodology behind the plots presented in Section 3, in providing the theoretical regions of the moment of skewness and excess kurtosis, is general, and should serve as a standard check of any new four-parameter distribution to provide information on the flexibility of the distribution in modeling skewness and kurtosis. The bucket plot can also be used to investigate how adequately or inadequately a new theoretical distribution fits a response variable, compared to the existing ones, for specific data sets.

In conclusion, the (single and multiple) bucket plot provides an additional residual diagnostic tool that can be used for fitted model checking, alongside other diagnostic tools.

SUPPLEMENTARY MATERIAL

The main functions to reproduce the bucket plots are available in the R packages gamlss and gamlss.ggplots. The bucket_simulation_example file presents the R code to reproduce the results given in Table 1 and Figure 2. The code to reproduce the results presented for the analysis of the London Bike-sharing data can be obtained by sending and email to the correspondent author. The London Bike-sharing data is available at https://www.kaggle.com/hmavrodiev/london-bike-sharing-dataset.

Funding

The first author received financial support from National Council for Scientific and Technological Development (CNPq), Brazil, Process number 310050/2019-7, and Pró-Reitoria de Pós-Graduação (PROPG) from Federal University of Pernambuco, Process number 23076.076712/2020-84.

References


NOLAN, J. P. (2012). *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/~jnolan.


Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife - PE, Brazil, 50670-901 E-mail: fernanda.bastiani@ufpe.br E-mail: d.stasinopoulos@londonmet.ac.uk E-mail: r.rigby@londonmet.ac.uk

Balaclava Rd, Macquarie Park NSW 2109, Austrália E-mail: gillian.heller@mq.edu.au