

***Short note***

**The Effect of Magnitude Uncertainty on Earthquake  
Activity Rates**

*RMW Musson*

*British Geological Survey*

*West Mains Road*

*Edinburgh EH9 3LA*

*UK*

[rmwm@bgs.ac.uk](mailto:rmwm@bgs.ac.uk)

*+44 131 650 0205*

## ***Abstract***

At present, any seismic hazard analyst seeking advice from the literature on how to handle uncertainty in magnitude values when calculating activity rates for seismic source zones may be alarmed to find two different viewpoints that apparently contradict one another, and that papers advocating one approach fail to mention the other, and vice versa. Superficially, it appears to be demonstrable that the uncertainty in earthquake magnitude either causes an overestimation of the true activity rate, or causes an underestimation. In this short note, it will be demonstrated that the resolution to the dichotomy depends not only on whether magnitude data have been converted, but also on how, a point not previously made. Various authors have proposed a correction factor to remedy the effect of uncertainty on activity rate, but if this is applied wrongly, the problem may be exacerbated. In practice, actual cases may be complex and difficult to resolve.

## ***Introduction***

One of the routine tasks in probabilistic seismic hazard analysis (PSHA) is calculating the activity rate for any seismic source. This value, essentially the number of events per year above a threshold magnitude value, serves to anchor the magnitude-frequency distribution expressed by a slope known as the b value. The method for calculating b is most commonly the maximum likelihood estimator, proposed by Aki (1965), and developed further by Wiechert (1980).

Ideally, calculations should take account of the fact that the magnitude values in any earthquake catalogue are imprecise. Even when magnitudes are determined from good-quality modern instrumental data, one might expect an uncertainty of the order

of at least 0.2 magnitude units, and for historical data this is likely to be higher. Estimation of the uncertainty, especially for historical events, is a complex issue that will not be addressed here.

So far as the literature is concerned, there is agreement that uncertainty in magnitudes does not affect the b value, but does impact on the activity rate. Unfortunately, the literature is somewhat confusing, due to a lack of cross-reference between some papers. The purpose of this short note is to highlight this issue and indicate an additional source of difficulty.

### ***Magnitude uncertainty***

The issue of uncertainty in magnitudes was first examined in 1985 by Veneziano and van Dyke (1985) in the USA, and Tinti and Mulargia (1985) in Italy. More recently, these results have been recapitulated by McGuire (2004) and Castellaro et al. (2006) and also studied in New Zealand by Rhoades (1996) and Rhoades and Dowrick (2000). According to the American authors, magnitude uncertainty principally results in an underestimate of activity rates. According to the Italian and New Zealand side, magnitude uncertainty principally results in an overestimate of activity rates. Neither refers to the other. It is rather important to get the issue right, as if one corrects value upwards to account for a perceived underestimate, and the original estimate is actually too high, the original error is compounded.

One can assume that for a given area, a number of earthquakes have occurred in the past, such that they could be described as a Gutenberg-Richter power law distribution in terms of moment magnitude,  $M_w$ , where  $M_w$  is assumed to relate directly to seismic moment release. For the purposes of PSHA, one wishes to know the number

of earthquakes above a threshold magnitude,  $M_w'$ . However, in most cases one cannot observe  $M_w$  directly. One has instead a substitute measure,  $M_x$ . In the case of McGuire (2004)  $M_x$  is epicentral intensity, but it could be log felt area, or a different magnitude scale (such as local or surface-wave magnitude).

McGuire's (2004) argument is roughly as follows: given an equation from which one can estimate  $M_w$  from  $M_x$ , there exists some threshold value  $M_x'$ , which, when converted, yields  $M_w'$ . Thus, if one ignores uncertainty, the number of events exceeding  $M_w'$  is simply the number of events exceeding  $M_x'$ . But in reality, some events  $< M_x'$  are truly  $> M_w'$  and vice versa. Therefore the number of events  $> M_w'$  is equal to the number of events  $> M_x'$  plus those  $< M_x'$  that are really  $> M_w'$ , minus those events  $> M_x'$  that are really  $< M_w'$ . However, because of the power law distribution of seismicity, there are always more events  $< M_x'$  than  $> M_x'$  (assuming uncertainty to follow a symmetrical distribution). Therefore the additional events that are  $< M_x' / > M_w'$  are significantly more than the number of events  $> M_x' / < M_w'$ . Therefore the true number of events  $> M_w'$  is significantly higher than just the number of events  $> M_x'$ . Therefore ignoring uncertainty causes the true activity rate to be underestimated.

The contrary argument runs as follows. Consider an earthquake of given magnitude  $M_w$ . It is not the case that all earthquakes of the same  $M_w$  magnitude will be represented in the catalogue by identical values of  $M_x$ . In the first case, there will be inherent fluctuations in the physical processes that lead from  $M_w$  to  $M_x$ , and in the second case,  $M_x$  is subject to measurement error. Thus an event with true magnitude  $< M_w'$  may be represented in the earthquake catalogue by a value  $> M_x'$ , and likewise an event  $> M_w'$  may appear as a value  $< M_x'$ . But once again (and again assuming

uncertainty to be symmetrical), the power-law distribution ensures that the first case is more common than the second, simply because there are more earthquakes  $<M_w'$  than  $>M_w'$ . Therefore there are more small events that are overestimated than large events that are underestimated. Therefore ignoring uncertainty causes the true activity rate to be overestimated.

According to Tinti and Mulargia (1985) and Castellaro et al. (2006) the shift in activity rate,  $\delta a$ , can be found from

$$\delta a = ( b^2 \sigma_m^2 ) / ( 2 \log_{10} e ) \quad (1)$$

where  $\sigma_m$  is the magnitude uncertainty. From this, one can compute a correction factor

$$m' = \delta a / b \quad (2)$$

as was found by Rhoades and Dowrick (2000). The presentation of these equations differs according to whether one uses  $b$  or  $\beta$ , where  $\beta = b \ln(10)$ . As shown by simulation results in Rhoades (1996), subtracting this correction factor from the magnitude values given in an earthquake catalogue allows the correct activity rate to be retrieved. The correction factor proposed by Veneziano and van Dyke (1985) is effectively the same value, but with varying sign. They propose that when uncertainty arises from direct estimation of  $M_w$  (i.e., from seismograms), the correction factor should be subtracted, but when  $M_w$  is estimated from some other parameter  $M_x$ , it should be added. This corrected magnitude value is referred to as  $M^*$ . It should not be thought that these corrected  $M^*$  values are to be published in place of the  $M_w$  values; they are temporary values used only within the activity rate calculations.

Assuming for simplicity that

$$M_x = M_w \pm \sigma \quad (3)$$

one can compute a synthetic catalogue in which the  $M_w$  values follow a perfect Gutenberg-Richter distribution, and the corresponding  $M_x$  values for each event. It will readily be found that the number of events exceeding a threshold value of interest is significantly higher when one looks at the  $M_x$  values compared to the  $M_w$  values. Therefore, if one knew equation (3), converting back from  $M_x$  to  $M_w$  would overestimate the activity as in equation (1). Since one does not know equation (3), it is necessary to obtain a regression for  $M_w$  from  $M_x$ ,

$$M_w = f [ M_x ] \quad (4)$$

and the effect of the power-law distribution is to bias the regression; the result is that the number of events with estimated  $M_w$  above the threshold value is now too low, and the activity is underestimated as in McGuire (2004).

There are complications when one attempts to put this scheme into practice. A difficulty arises with the  $M_x$  to  $M_w$  conversion. Typically, one has only a subset of events with both  $M_x$  and  $M_w$  values, and both  $M_x$  (observed) and  $M_w$  (observed) are subject to errors. The result obtained for equation (4) will be affected both by the properties of the subset of the catalogue available for regression (for instance, how closely the subset follows a power law distribution), and also the regression measures employed, as for instance least squares or orthogonal regression.

Different regression methods for earthquake magnitudes are discussed by Stromeyer et al. (2004), Castellaro et al. (2006), Castellaro and Bormann (2007), Gutdeutsch et al. (2011) and Lolli and Gasperini (2012) amongst others. The consensus opinion of these studies is that general orthogonal regression (GOR) is to be preferred over

ordinary least squares regression (OLS) for converting magnitudes. Castellaro et al. (2006) argue that least squares regression is only mathematically valid given certain assumptions: notably that the errors on the independent variable are at least an order of magnitude less than those on the dependent variable, which is manifestly not true in the case of earthquake magnitude values. They also show from simulations that converting magnitudes using OLS leads to errors in estimating the  $b$  value for the Gutenberg-Richter relationship for the earthquake population.

If the ratio of magnitude variances cannot be computed, simple orthogonal regression (SOR) is preferable to OLS (Castellaro and Bormann 2007). Also called particular orthogonal regression, it is the case where one assumes the ratio to be 1.0.

As an example, the authors of the unified catalogue of earthquakes in central, northern, and northwestern Europe (CENEC), converted a range of magnitudes to  $M_w$  using chi-square maximum likelihood regression (CSQ), an approach intended to pay special attention to the errors affecting the input data (Grünthal et al., 2009). Lolli and Gasperini (2012) have demonstrated that there is little practical difference between CSQ and GOR.

## ***Simulation***

For purposes of demonstration, a large number of synthetic catalogues were constructed, each of 100 years, with events in the magnitude range  $M_w$  2.0 to 7.0, fitting a perfect Gutenberg-Richter distribution with  $b$  value of 1.0 and an activity rate set so that on average, each catalogue contains 30 events of  $M_w$  4.0 or above. The interest in this analysis is in events of  $M_w$  4.0 and above, but it is necessary to simulate lower magnitudes to account for events that scatter across the  $M_w$  4.0

threshold in both directions. This is similar to the simulation exercise undertaken by Castellaro and Bormann (2007), but here the objective is not to examine the reliability of different methods in estimating the slope of the regression line, but to look at the effect on estimates of the activity rate obtained from converted magnitudes.

It is assumed that all events can be measured in terms of  $M_x$ , derived from equation (3), with  $\sigma = 0.4$ . If one knew equation (3), one could convert to  $M_w$  and apply the correction factor from equation (2) when calculating the activity rate. The problem, again, is that equation (3) is not known, and has to be estimated. It is therefore assumed that a value  $M_w(\text{obs})$  is also known, which has a measurement error of  $\pm 0.2$ . Since both variances are known, GOR can be used.

A further problem relates to which events are actually available for study. Since this exercise is intended to mimic a real situation, it is unlikely that the complete simulated data will be available. Initial tests showed that different truncation strategies, e.g., taking all events with  $M_x \geq 4.0$ , or all events with  $M_w(\text{obs}) \geq 4.0$ , biased the regression in different ways. To try and mimic real data sets, the assumption was made that all events with both  $M_x \geq 4.0$  and  $M_w(\text{obs}) \geq 4.0$  would definitely be recorded, all events with either  $M_x \leq 2.0$  or  $M_w(\text{obs}) \leq 2.0$  would definitely not be recorded, and other events would have a probability of being recorded  $P$ , where

$$P = 0.5 * \min [ M_x, M_w(\text{obs}) ] - 2.0 \quad (5)$$

Figure 1 shows an example of such a data set, with the computed conversions from  $M_x$  to  $M_w$  using OLS and GOR. These are:

$$M_w = 0.59 + 0.77 M_x \quad (6)$$



from OLS and

$$M_w = -0.24 + 1.04 M_x \quad (7)$$

from GOR. In this particular simulation, the true number of earthquakes with  $M_w \geq 4.0$  was 27; the estimated number using equation (6) was 19, and the estimated number using equation (7) was 40. (It will be noticed in passing that the slope of the regression in equation (7) is very close to the true value of 1.0). So for this case, one conversion strategy results in estimated  $M_w$  values that underestimate the true number, while the other strategy results in an overestimation.

The simulations were repeated 1,000 times. The mean percentage underestimation of the number of events with  $M_w \geq 4.0$  from conversion with OLS was  $36.11\% \pm 12.67\%$ , whereas the mean overestimation from conversion with GOR was  $32.31\% \pm 19.34\%$ . The full results are shown in Figures 2 and 3.

When the  $M^*$  correction is applied to magnitudes converted using OLS, the results improve greatly, and the mean error over 1,000 simulations is an overestimation of  $6.02\% \pm 15.52\%$ . But it is clear that applying the same correction to the values converted using GOR would produce a gross overestimation of the activity. Equation (7) is a good approximation to the inverse of equation (3), hence the true activity rate can be recovered by subtracting the conversion factor from equation (2); the mean error becomes an underestimate of  $8.00\% \pm 13.96\%$ . The distributions are shown in Figures 4 and 5.

## ***Discussion***

These results suggest that the effect of uncertainty due to magnitude conversion is unstable, and depends on the properties of the data set used for determining the equation for converting one magnitude to another and the regression method. In the simulations conducted here, following the procedures described by McGuire (2004) gives reasonable overall results with conversions using OLS, within the design of this experiment (perfect Gutenberg-Richter fit, tapered truncation of sample, etc). When GOR was used, the correction factor must be subtracted rather than added, as in cases where one is only dealing with uncertainty in native  $M_w$  magnitudes. The lesson for the hazard analyst is that applying a correction factor by rote, without careful consideration of the underlying data sources, is liable to lead to error.

The more the conversion equation  $M_x$  to  $M_w$  approximates to the inversion of the underlying relationship that determines  $M_x$  from  $M_w$  (e.g., equation 4 is the inversion of equation 3), the more the uncertainty becomes equivalent to the uncertainty on directly observed  $M_w$ . In such cases, subtracting the correction factor is appropriate, even though the uncertainty involves a magnitude conversion.

The issue is further complicated by the increasingly common use of more complex methods of estimating magnitude from macroseismic data, of which Bakun and Wentworth (1997) is one example of several. The calibration of such methods is complex and far from the simple regression between magnitude and epicentral intensity used as an example by McGuire (2004). Although one can obtain an estimate of uncertainty in magnitude from such methods, for instance by bootstrapping, it is far from clear what the impact of such values really is on the activity rate. There will

exist complex cases where magnitudes have been converted more than once, for instance, from macroseismic data to ML, and from ML to Mw.

When working with imported catalogs, for which one does not have a complete history of the derivation of each magnitude value, it can be that one knows that the values are uncertain, but not whether the activity rates are overestimated or underestimated.

A further problem is that, as can be seen in Figures 2 to 5, the simulations give a wide variety of results. In actual practice, one is generally dealing with a single earthquake catalog, which may fall anywhere in the distribution. So while, on balance of probability, applying a correction factor improves one's chances of accurately estimating the true activity rate, there is no guarantee that doing so will not result in a worse under- or overestimate.

### ***Conclusions***

1. When calculating earthquake activity rates from data that are an uncertain estimator of the desired magnitude measure, the true activity rate may be overestimated, or underestimated, depending on whether magnitudes are converted from another value, and (importantly) if so, how.
2. In cases where a single homogenous magnitude measure has been used throughout, the true activity rate can be recovered by reducing magnitude values before performing the activity rate calculations.
3. In cases where magnitudes are converted from some other parameter, success of the correction, and the direction in which it should be applied, depends on exactly how the conversion was performed, and this may be hard to assess.

Caution is therefore required in PSHA studies when attempting such corrections.

### ***Data and Resources***

None used.

### ***Acknowledgements***

This work has been partly supported by the European Commission through the project: FP7-ENVIRONMENT-226967 entitled “Seismic Hazard Harmonization in Europe”. The author would like to thank Hilmar Bungum, NORSAR, for comments and encouragement in writing this paper, Bob Youngs, AMEC Geomatrix, for helpful discussion, and David Kerridge, BGS, for reviewing the first draft. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC).

### ***References***

- Aki, K. (1965). Maximum likelihood estimation of  $b$  in the formula  $\log N = a - bM$  and its confidence limits. *Bull. of the Eq. Res. Inst., Tokyo University* **43**, 237-239.
- Castellaro, S., F. Mulargia, and Y. Y. Kagan. (2006). Regression problems for magnitudes. *Geophys. Jnl. Int.* **165**, 913-930.
- Castellaro, S., and P. Bormann. (2007). Performance of different regression procedures on the magnitude conversion problem. *Bull. Seis. Soc. Am.* **97**, 1167-1175.

- Grünthal, G., R. Wahlström, and D. Stromeyer. (2009). The unified catalogue of earthquakes in central, northern, and northwestern Europe (CENEC)—updated and expanded to the last millennium. *Jnl. Seismol.* **13** 4, 517-541.
- Gutdeutsch, R., S. Castellaro, and D. Kaiser. (2011). The magnitude conversion problem: Further insights. *Bull. Seis. Soc. Am.* **101**, 379-384.
- Lolli, B., and P. Gasperini. (2012). A comparison among general orthogonal regression methods applied to earthquake magnitude conversions. *Geophys. Jnl. Int.*, DOI: 10.1111/j.1365-246X.2012.05530.x.
- McGuire, R. K. (2004). *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute, Oakland CA. 221 pp.
- Rhoades, D. A. (1996). Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties. *Tectonophysics* **258**, 71-83.
- Rhoades, D. A., and D. J. Dowrick. (2000). Effects of magnitude uncertainties on seismic hazard estimates. *Proc. 12th World Conference on Earthquake Engineering, 30 Jan-4 Feb 2000, Wellington, NZ*. Paper 1179.
- Stromeyer, D., G. Grünthal, and R. Wahlström. (2004). Chi-square regression for seismic strength parameter relations, and their uncertainties, with applications to an Mw based earthquake catalogue for central, northern and northwestern Europe. *Jnl. Seismol.* **8**, 143-153.

Tinti, S., and F. Mulargia. (1985). Effects of magnitude uncertainties on estimating the parameters in the Gutenberg-Richter frequency-magnitude law. *Bull. Seis. Soc. Am.* **75**, 1681-1697.

Veneziano, D., and J. Van Dyke. (1985). Seismic parameter estimation methods. *in* EPRI/SOG, ed. *Seismic Hazard Methodology for Nuclear Facilities in the Eastern United States (Draft 85-1)*, volume 2, Appendix A.

Weichert, D. H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bull. Seis. Soc. Am.* **70**, 1337-1346.

*RMW Musson*

*British Geological Survey*

*West Mains Road*

*Edinburgh EH9 3LA*

*UK*

[rmwm@bgs.ac.uk](mailto:rmwm@bgs.ac.uk)

*+44 131 650 0205*

## **Figure captions**

### **Figure 1**

Example of regression data (from simulation) for deriving an  $M_x$  to  $M_w$  conversion. Both the OLS and GOR linear fits are plotted.

### **Figure 2**

Distribution of the percentage overestimation of the actual number of earthquakes  $M_w \geq 4.0$  when  $M_w$  is converted from  $M_x$  using OLS, over 1,000 simulations ( $\sigma M_x = 0.4$ ,  $\sigma M_w = 0.2$ ). Underestimation is shown as negative values on the X axis.

### **Figure 3**

Distribution of the percentage overestimation of the actual number of earthquakes  $M_w \geq 4.0$  when  $M_w$  is converted from  $M_x$  using GOR, over 1,000 simulations ( $\sigma M_x = 0.4$ ,  $\sigma M_w = 0.2$ ).

### **Figure 4**

Distribution of the percentage overestimation of the actual number of earthquakes  $M_w \geq 4.0$  when  $M_w$  is converted from  $M_x$  using OLS and adding a correction factor, over 1,000 simulations ( $\sigma M_x = 0.4$ ,  $\sigma M_w = 0.2$ ). The distribution shows that while the activity rates based on corrected magnitudes are a much better representation of the true values than using uncorrected magnitudes as in Figure 2, there is a chance that activity rates in an individual case will significantly overestimate the true values.

***Figure 5***

Distribution of the percentage overestimation of the actual number of earthquakes  $M_w \geq 4.0$  when  $M_w$  is converted from  $M_x$  using GOR and subtracting a correction factor, over 1,000 simulations ( $\sigma M_x = 0.4$ ,  $\sigma M_w = 0.2$ ).



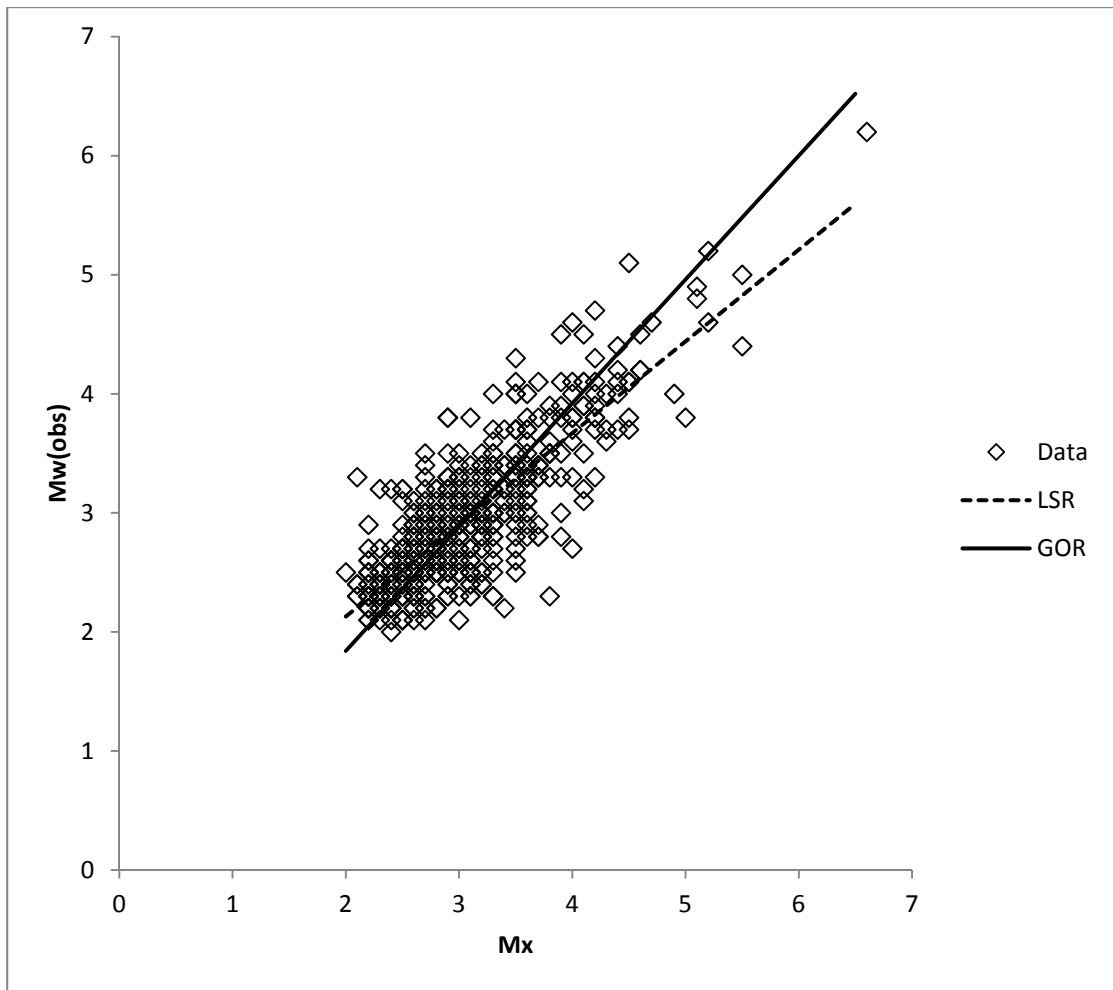


Figure 1

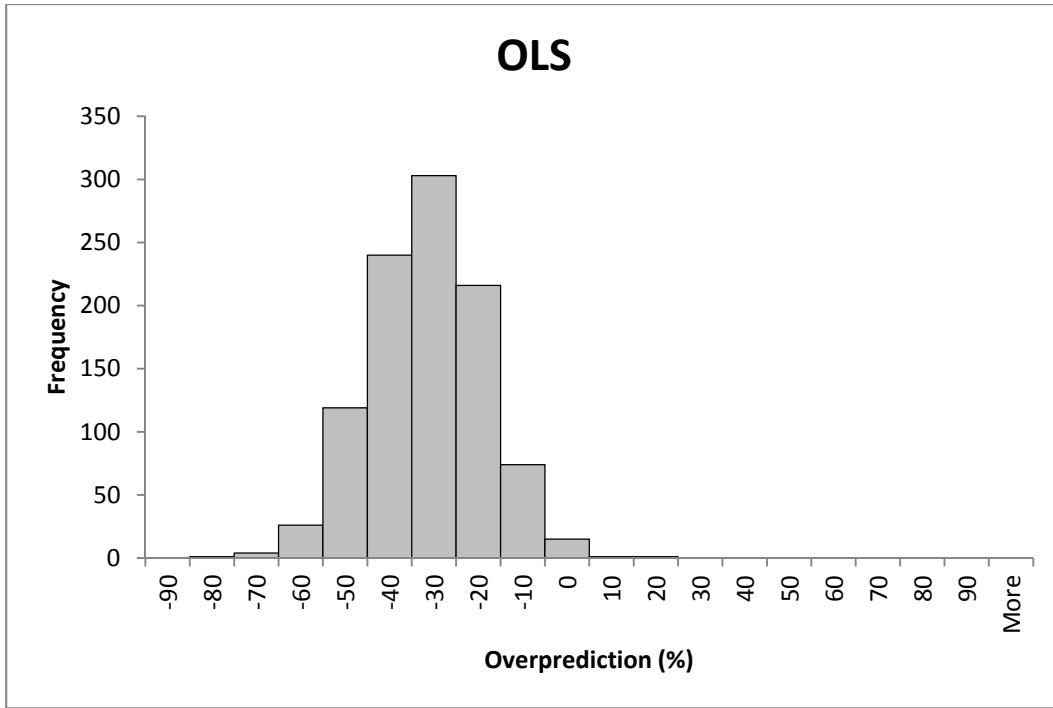


Figure 2

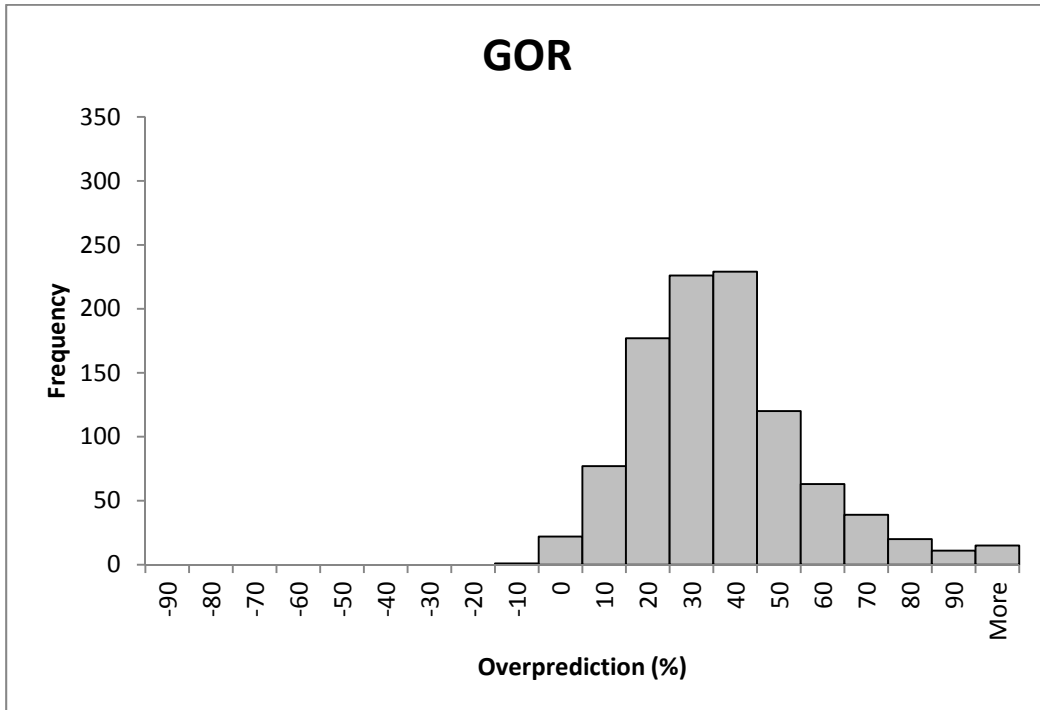


Figure 3

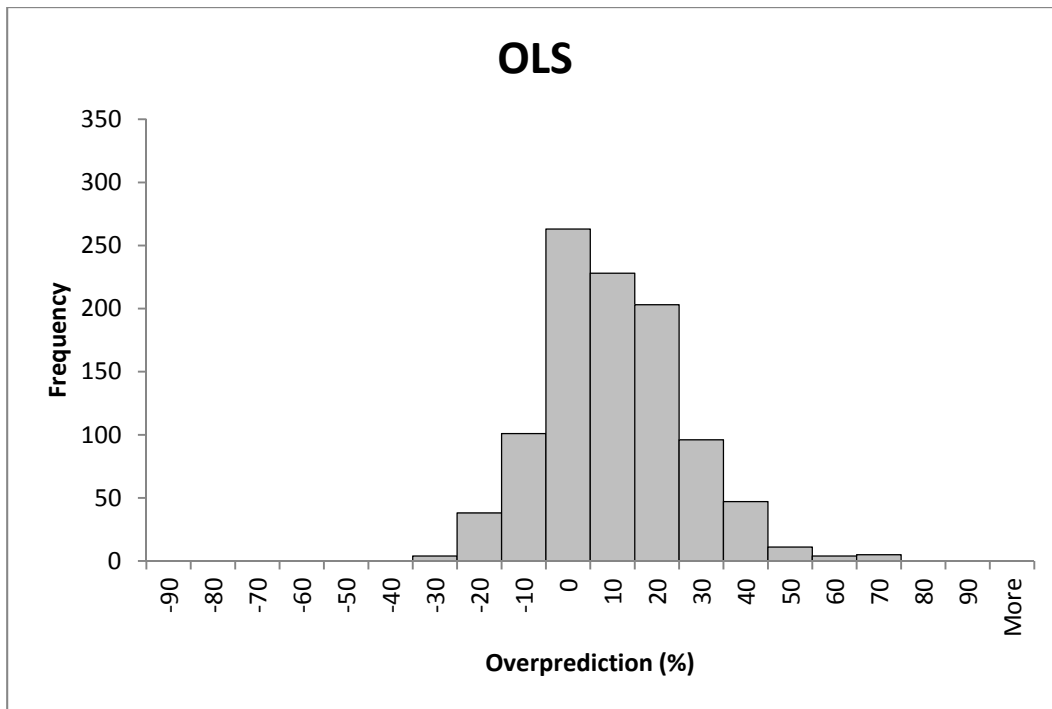


Figure 4

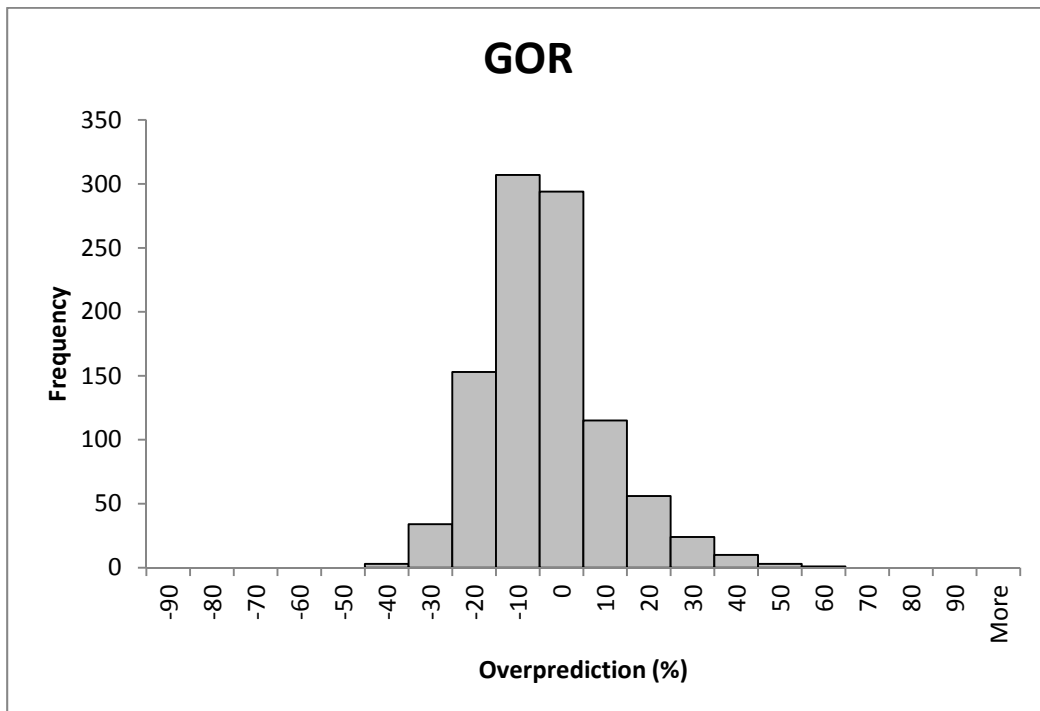


Figure 5