16th World Conference on Earthquake, 16WCEE 2017
Santiago Chile, January 9th to 13th 2017
Paper N° 2617
Registration Code: S-E1461961577

WHEN ADDRESSING EPISTEMIC UNCERTAINTY IN A LOGNORMAL FRAGILITY FUNCTION, HOW SHOULD ONE ADJUST THE MEDIAN?

K. Porter[(1)]

[(1)] *Research Professor, University of Colorado Boulder, keith.porter@colorado.edu*

Abstract

Probabilistic seismic risk analyses use fragility functions that relate the probability of an asset exceeding specified limit states to the seismic excitation to which the asset is subjected. The fragility function is commonly idealized with the two-parameter lognormal cumulative distribution function (CDF). One may wish to add uncertainty to such a fragility function to account for epistemic uncertainties. For example, one may want to recognize that the fragility function was derived from too-limited data that might inadequately represent the population of assets in question. Or one may want to counteract the decrease in uncertainty resulting from some other modeling simplification. Engineers commonly increase the uncertainty by taking the standard deviation of the natural logarithm of the capacity (the combined or total uncertainty) to be the square root of the sum of the squares of the initial uncertainty (often called aleatory) and epistemic uncertainty. Some engineers have recognized that, if one does not also increase the median, one tends to bias loss estimates high. They advocate rotating the fragility function about the 10th percentile. I found the available evidence supporting the use of the 10^{th} percentile unsatisfying, so I studied how different values of the rotation point might bias long-term failure rate. This work examines how the long-term rate is sensitive to the rotation point, median, aleatory and epistemic uncertainties, and site hazard. It appears better to rotate fragility functions about the 20th percentile if one is concerned with long-term failure rate and expected annualized loss. If one is more concerned with extrema, it can be better to rotate about the 10th or 50th percentiles.

Keywords: fragility; bias; probabilistic seismic risk

1. Problem: increasing uncertainty in a fragility function without introducing bias in risk

Like many probabilistic seismic risk analysis procedures, the Global Earthquake Model's analytical vulnerability guidelines [e.g., 1] make frequent use of lognormal cumulative distributions functions (CDFs) to idealize fragility functions, as in Eq. (1). In Porter et al. [1], we use lognormal CDFs to idealize damage to building components and the collapse of buildings, but lognormal CDFs are also commonly used to express the probability that a building, bridge, or other facility will exceed various other limit states. In the equation, P[A|B] denotes the probability that statement A is true given that statement B is true; D denotes the uncertain damage state of some particular asset; d is a particular value of D (a particular damage state); X denotes the uncertain excitation to which the asset is subjected; x denotes a particular value of that excitation; Φ denotes the Gaussian cumulate distribution function; ln denotes the natural logarithm of the quantity in parentheses; and θ and β are parameters of the distribution: θ denotes the median; β, the standard deviation of the natural logarithm of capacity, or more briefly the logarithmic standard deviation. An example lognormal fragility function is shown in Fig. 1. D is assumed here to be an ordinal number, meaning that if the asset can enter more than one damage state, the damage states can be ordered, so that damage state D = 2 is somehow worse than damage state D = 1.

$$P\left[D \geq d \middle| X = x\right] = \Phi\left(\frac{\ln\left(x/\theta\right)}{\beta}\right) \tag{1}$$

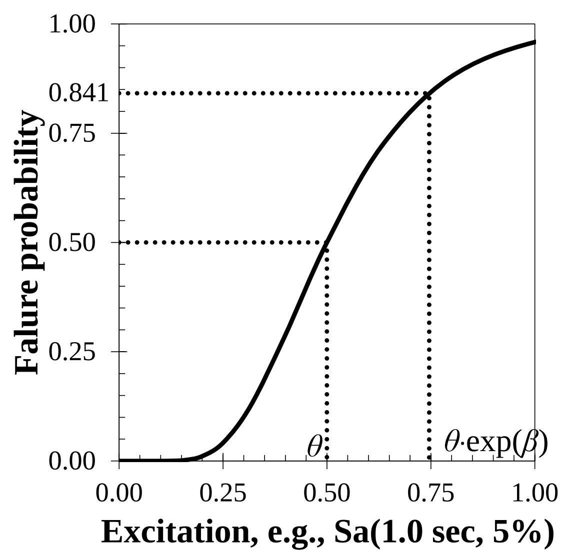


Fig. 1 – Example lognormal fragility function

The fragility parameters are often conditioned on or used in conjunction with some modeling simplification. In the case of the damage analysis of a component at a particular story in a particular building subjected to a particular base excitation, the simplification may be in the structural analysis, where we calculate the demand parameter to which the component is subjected. For example, we might calculate the peak transient drift ratio to which the component is subjected using a nonlinear pseudostatic structural analysis procedure. The base excitation might be measured in terms of the spectral acceleration response at some fixed period T and damping ratio z, and because of the structural analysis technique, the peak transient drift ratio in question has no associated uncertainty. The analyst knows that there would in fact be variability in drift conditioned on $S_a(T,z)$, because record-to-record variability among ground motions with the same $S_a(T,z)$ produces different drifts. How then to incorporate the added uncertainty?

2. Options

At least two options present themselves: do nothing, or increase the original β with an additional value to reflect modeling uncertainty. Kennedy and Short [2], NIBS and FEMA [3], and ATC [4] all increase uncertainty at various stages of probabilistic seismic risk analysis. Let β_u denote the increase in the logarithmic standard deviation of capacity. Let β_d denote the logarithmic standard deviation of capacity before increasing uncertainty.

The do-nothing alternative (β_u = 0) offers simplicity but opens the analysis to easy attacks on the basis of underestimated uncertainty. Including β_u avoids attacks on the basis of underestimated uncertainty, can be done with a fairly simple calculation, and makes the analysis consistent with common practice as suggested above. In [1], the author considered these two options, their advantages and disadvantages, and selected the latter, that is, increasing uncertainty. However, doing so raises the question of exactly how to include the added β.

The previously cited authors combined two logarithmic standard deviations together by calculating the square root of the sum of their squares (SRSS) and used that value as the total logarithmic standard deviations (usually, anyway; a more complicated convolution is required to reflect a lognormally distributed excitation). This is justified by the consideration that if the quantity in question, the uncertain capacity of a building or component, is taken as the product of two uncertain, lognormally distributed, uncorrelated quantities, then their logarithmic variances properly sum and equivalently their logarithmic standard deviations are SRSSd. Let us proceed then to SRSS the logarithmic standard deviations as prior authors have done, as shown in Equation (2).

$$\beta_c = \sqrt{\beta_r^2 + \beta_u^2} \tag{2}$$

Kennedy and Short [4] found that keeping θ constant tends to bias the mean failure rate, denoted here by λ. They found that failure rate is insensitive to β_c if one rotates about the 10th percentile. That is, λ is not greatly changed in one increases β by Equation (2) and increases θ so that the value of excitation x associated with failure probability of 0.10 is unchanged after increasing β and θ. Let p denote the y-value of the rotation point, i.e., the point shared by the pre- and post-rotation fragility functions. Then θ' is given by Equation (3). The concept of rotating around p is illustrated in Fig. **2**.

$$\theta' = \theta \cdot \exp\left(-\Phi^{-1}\left(p\right)\cdot\left(\beta_c - \beta_d\right)\right) \tag{3}$$

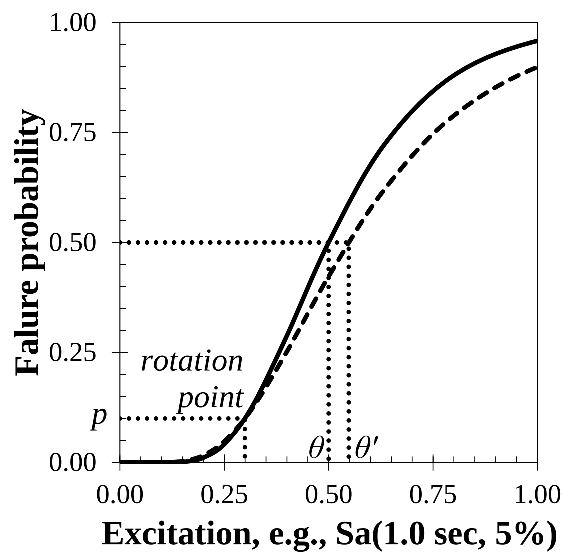


Fig. 2 – Illustration of rotating the fragility function about p

Let us denote the failure rate after rotating about p by λ', and let θ' denote the median capacity after rotating about p. Let $G(x)$ denote the expected value of the rate at which earthquakes occur at a particular site and cause shaking of at least x, in events per year. (Let us leave it ambiguous whether that is the rate of mainshocks or the rate of all earthquakes.) Then λ and λ' are given by Equations (4) and (5) respectively.

$$\lambda = \int_{x=0}^{\infty} \Phi\left(\frac{\ln\left(x/\theta\right)}{\beta_d}\right)\left|\frac{dG\left(x\right)}{dx}\right| dx \tag{4}$$

$$\lambda' = \int_{s=0}^{\infty} \Phi\left(\frac{\ln\left(x/\theta'\right)}{\beta_c}\right)\left|\frac{dG\left(x\right)}{dx}\right| dx \tag{5}$$

3. Tornado diagrams for sensitivity testing

Let us begin the assessment of where to rotate with a series of sensitivity tests depicted in the form of tornado diagrams. See Howard [5] for an early general discussion of tornado diagrams, or Porter et al. [6] for an early discussion of tornado diagrams applied to seismic risk. Briefly, a tornado diagram depicts the sensitivity of a dependent variable to each of two or more uncertain independent variables. One chooses baseline values of the independent variables, along with high and low values of each independent variable. One can select probability-related definitions of baseline, high and low, such as mean or median, 90th, and 10th percentiles respectively, but it is not required. The dependent variable is assessed with all independent variables set to their baseline values. This is the baseline value for the dependent variable.

Then the first dependent variable is set to its high value and the dependent variable evaluated. The first dependent variable is next set to its low value and the dependent variable evaluated again. The difference between these two quantities is referred to as the swing for the 1st dependent variable. The first dependent variable is set back to its baseline value and the process repeated for the 2nd independent variable. The swing for the 2nd independent variable is evaluated. The process is repeated for each remaining independent variable: all are set to their baseline values and only one independent variable is varied at a time, i.e., not simultaneously.

The independent variables are sorted in decreasing swing and the results plotted with a horizontal bar chart, one bar for each independent variable. The independent variable with the highest swing is shown with the top bar, the 2nd highest swing with the second highest bar, etc. The dependent variable is measured on the horizontal axis. The ends of the bars are placed at the value of the dependent variable resulting from the low and high values of the independent variable in question. A vertical line is drawn through the baseline value of the dependent variable (i.e., with all the independent variables set at their baseline values).

The result is a bar chart that resembles a tornado in profile, hence the name. The bars at the top of the chart are the ones to which the dependent variable is most sensitive; the ones at the bottom, the least. One can use the tornado diagram to better decide which independent variable to investigate more closely, or to try to reduce its uncertainty, or other uses.

4. Applying tornado diagrams to the present problem

For present purposes, we want to see how sensitive mean failure rate λ is to the rotation point p, and just to be sure p is important, let us also check sensitivity to median capacity θ, the logarithmic standard deviations β_d and β_u, and the seismic hazard of the location where the asset is placed. The rationale is that if λ is more sensitive to θ or β than to p, there is less reason to worry about the best value of p. As with Kennedy and Short's [2] example, let the dependent variable relate to the error in mean failure rate. Let us arbitrarily select 30 US locations with moderate to high seismic hazard, defined as a location where an ordinary building would have an ASCE 7-05 seismic design category B or above, i.e., $S_{D1} \geq 0.067$g. The locations are shown in Fig. 3. The figure shows locations and a shaded map where the three shades from dark to light are high, moderate, and low seismicity, defined using the breakpoints S_{D1} = 0.20g and 0.067g. (That is, high would be a count where at least one point has an ordinary building would have SDC D or above, moderate has at least one location with SDC B or C, and low is SDC A.) They were selected to reflect a variety of hazard levels and tectonic environments (plate boundaries, continental interior, generally strike-slip, generally subduction).

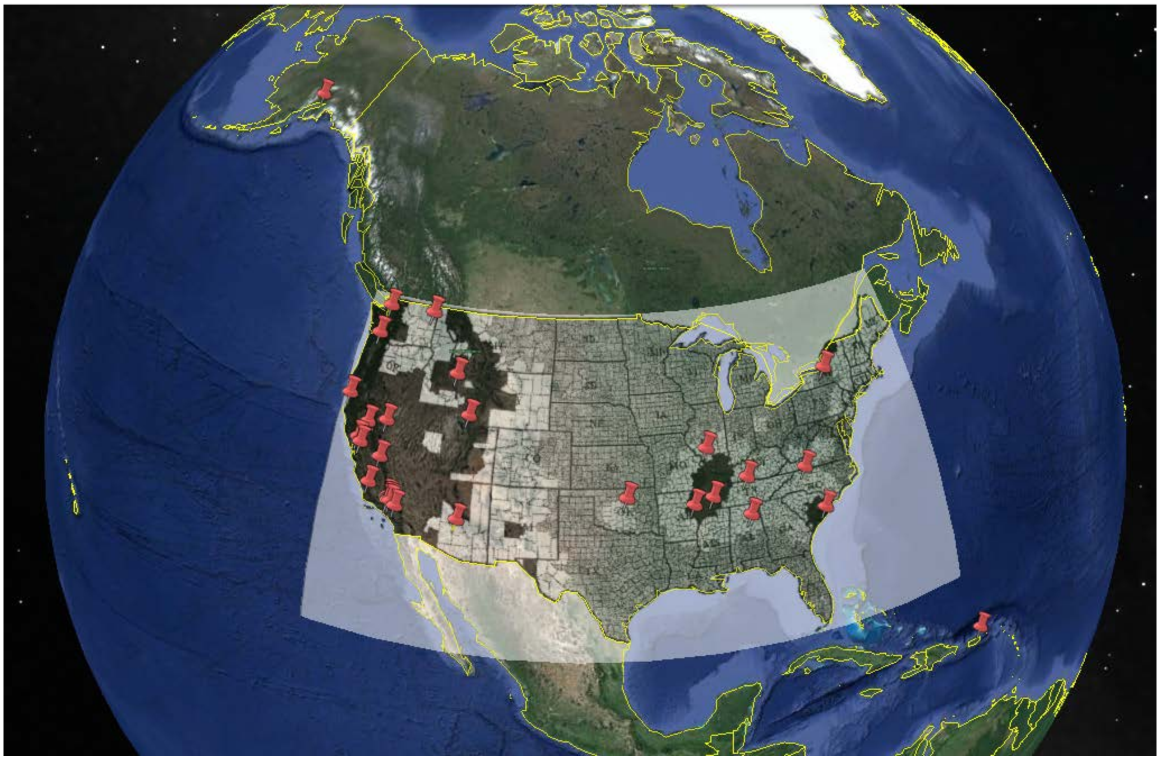


Fig. 3 – Sample locations

Let us consider several error measures:

(1) Average absolute error, averaging error over 30 US locations with moderate to high seismic hazard. By "absolute error" is meant the difference between λ with rotation and λ without rotation. By "average" is meant that we average the difference over the n = 30 locations. That is,

$$\varepsilon_{aa} = \frac{1}{n}\sum_{i=1}^{n}\left(\lambda_i' - \lambda_i\right) \tag{6}$$

(2) Average relative error, same locations. By "relative error" is meant the absolute error divided by the λ without rotation. By "average" is meant the average of the relative error over the 30 locations.

$$\varepsilon_{ar} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lambda_i' - \lambda_i}{\lambda_i}\right) \tag{7}$$

(3) Maximum absolute error, max over the 30 locations. By "maximum" is meant the largest of the 30 absolute values of the difference between λs with and without rotation. That is,

$$\varepsilon_{xa} = \max_i\left(\lambda_i' - \lambda_i\right) \tag{8}$$

(4) Minimum absolute error, min over 30 locations. Similar, but minimum. That is,

$$\varepsilon_{na} = \min_i\left(\lambda_i' - \lambda_i\right) \tag{9}$$

(5) Maximum relative error, max over 30 locations. Like 3, but normalized by λ without rotation.

$$\varepsilon_{xr} = \max_i\left(\frac{\lambda_i' - \lambda_i}{\lambda_i}\right) \tag{10}$$

(6) Minimum relative error, min over 30 locations. Like 4, but normalized by λ without rotation.

$$\varepsilon_{nr} = \min_i \left(\frac{\lambda_i' - \lambda_i}{\lambda_i} \right) \tag{11}$$

(7) Relative error at the location of highest hazard. Highest hazard measured in terms of ASCE 7-05 S_{D1}.

$$\varepsilon_{hr} = \frac{\lambda_i' - \lambda_i}{\lambda_i} \qquad i : \max_i \left(S_{D1,i} \right) \tag{12}$$

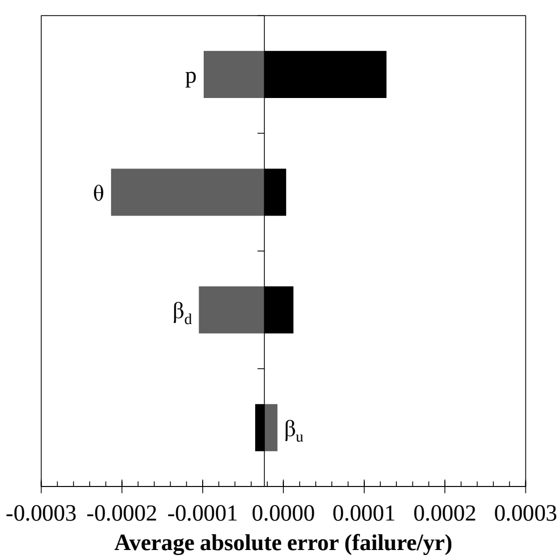
(8) Relative error at the location of lowest hazard. Lowest hazard in terms of ASCE 7-05 S_{D1}.

$$\varepsilon_{lr} = \frac{\lambda_i' - \lambda_i}{\lambda_i} \qquad i : \min_i \left(S_{D1,i} \right) \tag{13}$$

Baseline, low, and high values are as shown in Table 1. Tornado diagrams for these 8 dependent variables follow the table.

Table 1 –Sensitivity test parameter values

	Trial values		
	Low	Baseline	High
median, θ	0.2	0.5	1.0
logstdev without rotation, β_d	0.2	0.4	0.6
logstdev to reflect uncertainty, β_u	0.1	0.2	0.3
rotation point (percentile), p	0.1	0.2	0.5



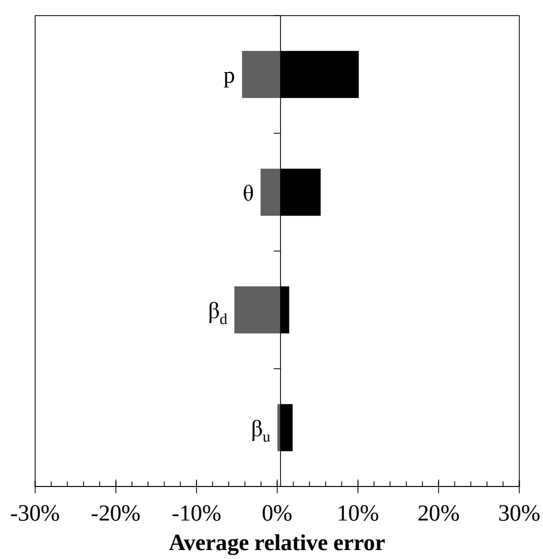


Fig. 4A – Tornado diagrams depicting sensitivity of average absolute error and average relative error to fragility parameters. Gray reflects the lower value of the independent variable; black, the higher value.

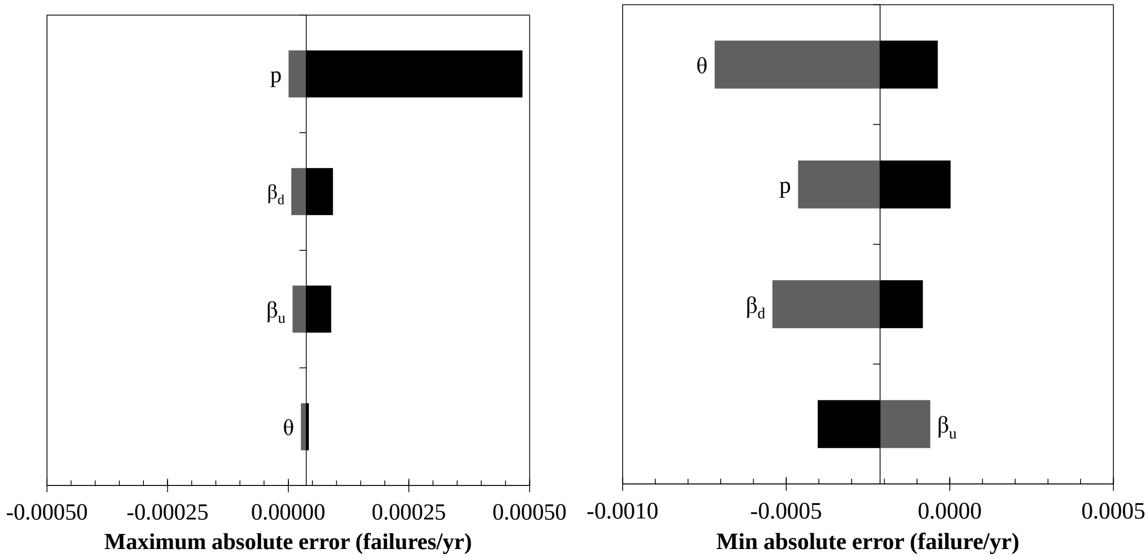


Fig. 4B – Tornado diagrams depicting sensitivity of máximum and mínimum absolute error to fragility parameters. Gray reflects the lower value of the independent variable; black, the higher value.

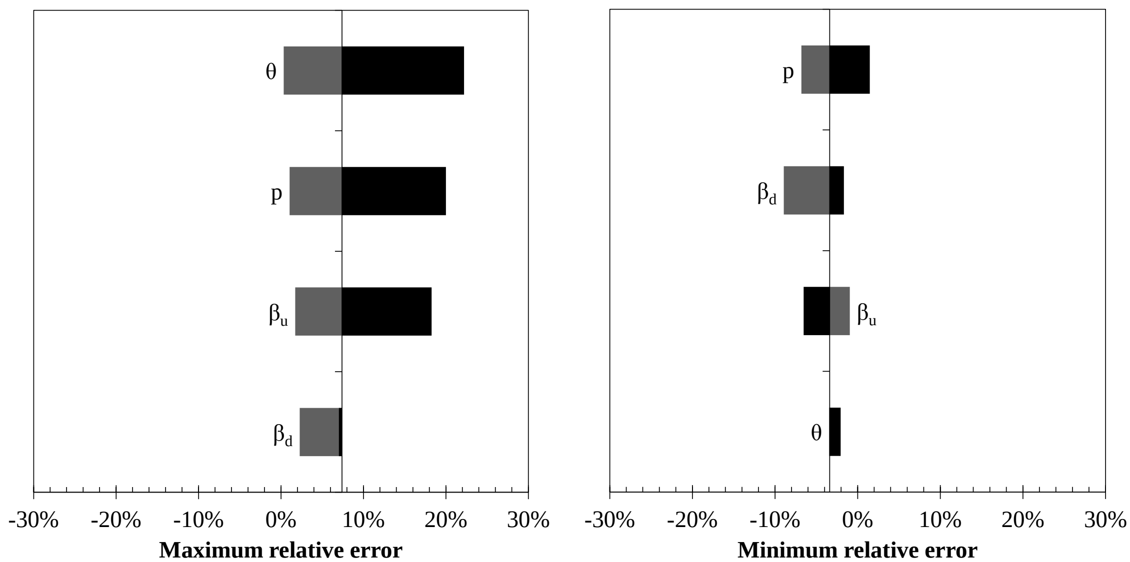


Fig. 4C – Tornado diagrams depicting sensitivity of máximum and mínimum relative error to fragility parameters. Gray reflects the lower value of the independent variable; black, the higher value.

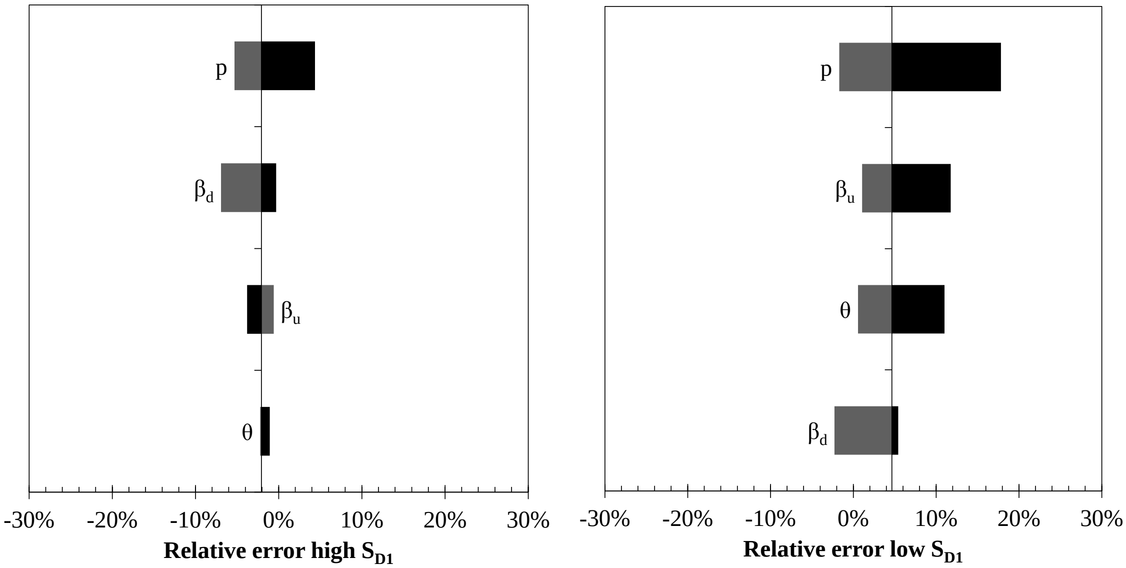


Fig. 4D – Tornado diagrams depicting sensitivity of relate error in high and low SD1 to fragility parameters. Gray reflects the lower value of the independent variable; black, the higher value.

5. Observations from the tornado diagrams

First, the rotation point *p* is always the top, sometimes the second, most important variable, which implies that it is worth worrying about the selection of *p*. That is, *p* matters. It is the top variable for the average relative error (smallest $|\varepsilon_{ar}|$), which seem to be the most relevant one for a portfolio probabilistic seismic risk analysis. The second-highest variable is generally though not always θ.

Second, continuing with the assumption that we are primarily concerned about the average relative error ε_{ar}, p = 0.20 is the best choice of the three: the baseline value (p = 0.20) is the closest of the three values to producing ε = 0.00, i.e., it produces the least average bias. It is also the best choice for high-hazard sites (smallest $|\varepsilon_{hr}|$) though not for low-hazard sites (smallest $|\varepsilon_{lr}|$). For a range of values of θ, p = 0.20 is a good choice, generally producing average relative error $|\varepsilon_{ar}| \leq 0.06$.

Three error measures favor p = 0.10: maximum absolute error and maximum relative error (smallest $|\varepsilon_{xa}|$ and smallest $|\varepsilon_{xr}|$), and relative error for the lowest hazard site (smallest $|\varepsilon_{lr}|$). Two favor p = 0.50: minimum absolute and minimum relative error (smallest $|\varepsilon_{na}|$ and smallest $|\varepsilon_{nr}|$). These five error measures are relevant if we are concerned with the minimizing error in extremes—the one site out of a portfolio with the highest or lowest hazard, or the one site where the rotation makes the biggest difference. Thus, for single-site risk analyses one might want to rotate about p = 0.10 or possibly 0.50. For general-purpose probabilistic seismic risk analysis however, these sensitivity tests suggest p = 0.20 is the best choice.

6. Honoring empirical fragility data

There is another consideration that has to do with honoring the data used to derive empirical fragility functions. Often one employs such fragility functions in probabilistic seismic risk analysis. Porter et al. [7] offer empirical seismic performance data regarding approximately 1,500 mechanical, electrical, and plumbing components shaken by earthquakes around the world between 1971 and 1999. The data come from the Electric Power Research Institute's eSQUG database [8]. Porter et al. [7] present the performance data in scatter diagrams where *x* denotes seismic excitation (peak ground acceleration) and *y* denotes failure probability. The scatter diagrams also include fragility functions derived from the data. The scatter diagrams show that the data tend to lie at *x* values where the *y*-value of the fragility function is between 0.0 and 0.20, i.e., at low failure rates.

This is intuitive because we like to design buildings and building components so that in real earthquakes they tend to have low failure rates. In any case, because the fragility functions are fit to the data, and because the data tend to lie at low failure probabilities, it is probably better to keep the fragility function fairly constant at those low failure probabilities. Adding uncertainty and rotating about the 50th percentile increases the fragility function at x-values below the median, meaning that it raises the fragility function across the range of x where the data lie. Rotating about low percentiles will tend to honor the data of empirically derived fragility functions better than rotating about the 50th percentile. There tends to be little difference between fragility functions rotated about the 10th and 20th percentiles, so the choice between the two does not matter much compared with choosing between them and the 50th.

7. Other considerations

As previously noted we tend to design buildings and other assets not to fail in common earthquakes. "Common earthquakes" here mean those earthquakes that produce something near the average failure rate in some arbitrary asset, such as building collapses, where the averaging is over individual earthquakes. This means that when earthquakes do shake assets, failure rates are more commonly low than high. This is not always the case, but in general it tends to be true. Rotating about the 10th or 20th percentiles means that we will tend not to overestimate failure rates in common earthquakes; rotating about the 50th will tend to overestimate failure rates in common earthquakes, but be more accurate in rare earthquakes in which on the order of half or more assets fail.

8. Conclusions

It appears best to rotate fragility functions about the 20th percentile, especially in probabilistic seismic risk analysis of asset portfolios. If one is more concerned with extrema, it can be better to rotate about the 10th or 50th. This conclusion is based on sensitivity tests that show the least bias in average error considering a variety of locations, plus considerations of honoring empirical fragility data and minimizing bias in common earthquakes.

9. References cited

[1] Porter K, Farokhnia K, Vamvatksikos D, Cho I (2015): *Guidelines for component-based analytical vulnerability assessment of buildings and nonstructural elements*, Global Vulnerability Consortium, www.sparisk.com/pubs/Porter-2015-GEM-Analytical-Vulnerability.pdf

[2] Kennedy RP, Short SA (1994): *Basis for Seismic Provisions of DOE-STD-1020.* Prepared for US Department of Energy under contract to Lawrence Livermore National Laboratory and Brookhaven National Laboratory. National Technical Information Service, US Department of Commerce, Springfield, VA., 65 pp.

[3] (NIBS and FEMA) National Institute of Building Sciences and Federal Emergency Management Agency (2009): *Multi-hazard Loss Estimation Methodology, Earthquake Model, HAZUS®MH MR4 Technical Manual.* Federal Emergency Management Agency, Washington, DC

[4] Applied Technology Council (2012): *Guidelines for Seismic Performance Assessment of Buildings, FEMA P-58*, Redwood City, CA.

[5] Howard RA (1988): Decision analysis: practice and promise. *Management Science* 34 (6), 679-695

[6] Porter KA, Beck JL, Shaikhutdinov RV (2002): Sensitivity of building loss estimates to major uncertain variables. *Earthquake Spectra*, 18 (4), 719-743

[7] Porter KA, Johnson G, Sheppard R, Bachman R (2010): Fragility of mechanical, electrical, and plumbing equipment. *Earthquake Spectra* 26 (2) 451-472, http://www.sparisk.com/pubs/Porter-2010-MEP-fragility-1.pdf

[8] Electric Power Research Institute (2003): *SQUG Electronic Database User's Guide, eSQUG, Version 2.0 (QA)—Revision to TR-113705*, Palo Alto, CA, http://www.epriweb.com/public/RS_1009040.pdf (viewed 1 May 2008).