Structural Safety 96 (2022) 102185

Contents lists available at ScienceDirect

Structural Safety

journal homepage: www.elsevier.com/locate/strusafe

Quantifying the effect of probability model misspecification in seismic collapse risk assessment

Laxman Dahal [a,*], Henry Burton [a], Samuel Onyambu [b]

[a] Department of Civil and Environmental Engineering, University of California Los Angeles, CA, 90095, USA
[b] Department of Statistics University of California Los Angeles, CA, 90095, USA

ARTICLE INFO

Keywords:
Model misspecification
Sandwich estimators
Collapse uncertainty
Fisher information equality test

ABSTRACT

One of the main steps in probabilistic seismic collapse risk assessment is estimating the fragility function parameters. The maximum likelihood estimation (MLE) approach, which is widely used for this purpose, contains the underlying assumption that the likelihood function is known to follow a specified parametric probability distribution. However, this assumed distribution may not always be consistent with the "true" probability distribution of the collapse data. This paper implements the Information matrix equivalence theorem to identify the presence of model misspecification i.e., if the assumed collapse probability distribution is, in fact, the "true" one. In the presence of model misspecification, the fragility parameter estimates continue to be asymptotically normally distributed but the variance–covariance matrix is no longer equal to the inverse of the Fisher's Information matrix. To increase the robustness of the variance–covariance matrix, the Huber–White sandwich estimator is implemented. Using collapse data from eight woodframe buildings, the effect of model misspecification on fragility parameter estimates and collapse rate is quantified. For the considered building cases, the parameter estimation uncertainty in the collapse risk did not increase when the "sandwich" estimator was used compared to when probability model misspecification was not considered (i.e., using MLE). The proposed framework should be used to further investigate the issue of probability model misspecification as it relates to fragility parameter estimation since only a single construction type (woodframe buildings) and limit state (collapse) was considered in the current study.

1. Introduction

Fragility curves are widely used in modern performance-based earthquake engineering (PBEE) methodologies to quantify probabilistic seismic risk (e.g., [1]). Component level fragility functions describe the probability of exceeding a damage state of interest conditioned on a response demand (e.g., story drift, peak floor acceleration) level. At the building level, seismic fragility functions relate a ground motion intensity measure (IM) to a limit state of interest (e.g., collapse, post-earthquake structural safety). At either scale, seismic fragility functions are developed using empirical data from experiments or field observations or data generated from numerical analyses. Regardless of the data generation process, one of several alternative statistical procedures is used to fit an appropriate model. In general, fragility curves can be developed using either parametric or non-parametric approaches. Parametric fragility curves are based on a predefined functional form or probability distribution [2–5]. On the other hand, non-parametric fragilities do not need to conform to a specific functional form or distribution [6].

Collapse risk assessment is a central part of the 2nd generation PBEE procedure. Within this context, fragilities functions are used to quantify the probability of collapse conditioned on the IM level. By integrating the fragility function with an appropriate hazard curve, seismic collapse risk can then be assessed in terms of the mean annual frequency of exceedance or the probability of collapse over some predefined period (e.g., the service life of a building). The lognormal cumulative distribution function (CDF) is commonly used to fit parametric collapse fragilities [2–4,7]. The required data can be generated using procedures such as incremental dynamic analyses (IDAs) [8] or multiple stripe analysis (MSA) [9]. The IDA procedure incrementally scales each ground motion within a given record set until it causes collapse. The IM level at which collapse occurs for each ground motion is the dataset of interest. The method of moments [7] is then used to obtain the associated lognormal distribution parameters (θ, β), where θ defines the median collapse capacity while β is the dispersion or logarithmic standard deviation. The associated CDF is taken as the fragility function. MSA analyzes the structure at pre-defined hazard levels, which

* Corresponding author.
E-mail address: laxman.dahal@ucla.edu (L. Dahal).

https://doi.org/10.1016/j.strusafe.2022.102185
Received 23 August 2021; Received in revised form 29 November 2021; Accepted 3 January 2022
Available online 21 January 2022

are also used as the basis for the ground motion selection. In other words, unlike IDAs, a different suite of motions is selected and used in the analyses at each hazard level. Also, a binary outcome variable (e.g., 0 for non-collapse and 1 for collapse) corresponding to each ground motion and hazard level forms the basis of the MSA collapse data. Maximum likelihood estimation (MLE) is then used to obtain the lognormal distribution parameters that define the collapse fragility. The generalized linear model (GLM) with a Probit link function has been presented as an alternative to the MLE-based approach to estimating fragility parameters [2,3]. However, it is worth noting that the GLM approach also relies on maximization of the likelihood function. The primary difference between the two procedures is in the definition of the likelihood function. For both methods, a key assumption is that the random variable that describes the number of collapses at each hazard level follows a binomial distribution. In other words, the probability of a certain number of ground motion record(s) causing collapse at a given hazard level is assumed to follow a binomial distribution. This assumption is often made as a matter of mathematical convenience and its implication to the fragility parameter estimates and the uncertainty around those estimates has never been studied.

Within the PBEE framework, several sources of uncertainty are recognized and, to varying degrees, considered. In collapse risks assessments, record-to-record and modeling uncertainties have been the primary focus of prior research and practice. The former is much more commonly considered by utilizing one or more suites of ground motions for response history analysis [10]. Modeling uncertainty arises from the limitation in the data and knowledge that is used to develop the numerical model. Bradley [11] describes four categories of modeling uncertainty denoted as Types I through IV. Type I uncertainty is attributed to a lack of knowledge about the measured values of physical quantities or "basic" parameters (e.g., material strength and modulus). Errors in the models used to relate measured quantities to constitutive models constitute Type II uncertainty. Type III and Type IV modeling uncertainty are related to the chosen constitutive model and system-level idealization. Of the four categories, Types I and II are more commonly addressed in the research literature on seismic collapse risk assessments [12–17].

A third source of uncertainty that very few studies have addressed is associated with the parameter estimation procedure used to develop fragility functions (denoted as parameter estimation uncertainty in the remainder of the paper). Building upon previous work, Lallemant and Kiremidjian [18] derived closed-form confidence intervals for fragility curves based on asymptotic normality. More specifically, uncertainties in the GLM parameter estimates were quantified using the variance–covariance (herein referred to as covariance) matrix of the estimated parameters, which is computed by taking the inverse of the Fisher's information matrix. It is worth noting that this inverse relationship is only valid when the probability model that is attributed to the data is correctly specified [19]. Iervolino [20] quantified the uncertainty in seismic collapse rates using the Delta method, which is an approximate technique that uses a Taylor series expansion [21]. Both studies [18,20] assume the correct specification of the probability model for the collapse data. However, the validity of this assumption and the consequence of model misspecification have not been addressed in the literature in the context of building seismic collapse risk assessments. Other studies [22,23] have implemented various sampling (e.g., Monte Carlo simulation) and re-sampling (e.g., bootstrapping) techniques as a practical approach to quantifying the record-to-record and modeling uncertainty in the collapse rate. While sampling is a pragmatic tool to quantify uncertainty, it does not directly address the model misspecification issue. In fact, in the presence of model misspecification, the random samples drawn using the covariance obtained from the Fisher's information matrix is incorrect.

This paper focuses on the effect of probability model misspecification on collapse fragility parameter estimation and the associated uncertainty. The quasi-maximum likelihood estimator (QMLE) is used to fit collapse fragility functions and quantify the uncertainty in the associated lognormal parameters. We also utilize the fact that QMLE is generally a consistent parameter estimator which minimizes Kullback–Leibler Information Criterion (KLIC) [24]. Also known as KL divergence, the KLIC characterizes how far apart the true probability distribution is from the one that is assumed. In other words, minimizing KL divergence ensures that ignorance about the true probability model is minimized. QMLE is sometimes also referred to as the "minimum ignorance" estimator [25]. The uncertainty in the collapse fragility parameters obtained using QMLE is compared to the case where the correct probability model is assumed i.e., using the MLE approach. The next section of the paper outlines the theoretical properties of QMLE and sandwich estimators. Second, MSA collapse data from four unique single-family wood-frame buildings are used to implement the Information matrix equivalence theorem to identify for misspecification of a parametric model. Each building is analyzed for an existing and retrofitted condition resulting in collapse data for a total of eight buildings. MSA is conducted at 16 different intensity levels with 45 ground motions at each intensity [26]. Collapse risk including the associated parameter estimation uncertainty is quantified using both the MLE and QMLE procedures.

2. Huber–White (Robust) Sandwich estimators

Maximum likelihood is one of the most popular and principled tools used for estimation and inference [19,27]. Maximum likelihood relies on a fundamental assumption to ensure its key properties such as consistency [28] and asymptotic normality [29] hold for a broad range of applications. A parameter estimate (for instance, $\hat{\theta}$) is said to be consistent if $\hat{\theta} \rightarrow \theta^0$ as the sample size increases, where θ^0 is the "true" unknown parameter. Similarly, $\hat{\theta}$ is said to be asymptotically normal if $\sqrt{n}(\hat{\theta} - \theta^0)$ converges to a normal distribution with an asymptotic variance of the estimate i.e. $\sqrt{n}(\hat{\theta} - \theta^0) \rightarrow \mathrm{N}(0, \sigma^2_{\theta^0})$. The assumption in MLE is that the stochastic law which determines the behavior of the phenomena investigated (the "true" probability model) is known to lie within a specified parametric family of probability distributions (herein referred to as the model). In other words, the probability model is assumed to be "correctly specified". With respect to MLE-based parameter estimation in seismic collapse risk assessments, the lognormal probability distribution is assumed to govern the conditional probability of collapse. In many (if not most) circumstances, there may not be complete confidence in the validity of the assumption. The "Huber Sandwich Estimator" (herein referred to as sandwich estimators) is a framework used to estimate the variance of the QMLE (or parameter estimation uncertainty). When the underlying probability model is incorrect, pseudo-likelihood is being maximized instead of the true likelihood. The estimators are also commonly referred to as pseudo-maximum likelihood estimators or simply QMLE. White [25] showed that QMLE converges to a well-defined limit and meets the important properties of MLE such as consistency and asymptotic normality [27–29]. It is also shown that, with misspecification, the asymptotic variance matrix of the QMLE no longer equals the inverse of Fisher's information matrix. It is an important finding because being able to correctly specify the variance matrix is pivotal to data-driven function-fitting approaches. For instance, in parametric bootstrapping, the correctness of the parameters has a direct impact on the accuracy of the result.

When fitting a fragility, the likelihood function defines the probability that a random ground motion record causes collapse at a given IM level. The likelihood is assumed to follow a lognormal probability distribution function. This assumption is sufficient to estimate the mean values of the lognormal distribution parameters (θ, β). However, the assumption that the probability model is correctly specified can influence the parameter estimation uncertainty.

3. Parameter estimation uncertainty in collapse fragility functions

3.1. Maximum likelihood method

A fragility function defines the conditional probability of collapse as a function of the IM level. It can also be extended to other limit states such as different damage stages but only collapse is considered in the current study. The conditional probability of collapse is described by a lognormal CDF as shown in Eq. (1).

$$P(C|IM = im) = \Phi\left(\frac{\ln(im/\theta)}{\beta}\right) \tag{1}$$

where $P(C|IM = im)$ is the probability of collapse conditioned on $IM = im$, θ and β are the median and standard deviation of $\ln(IM)$, respectively, and Φ is the CDF of the standard normal distribution. The parameters that define the fragility function are θ and β. Various statistical procedures can be used to obtain the parameter estimates $\hat{\theta}$ and $\hat{\beta}$. A primary goal of the current study is to compare the central tendency values and associated uncertainty when MLE and QMLE are used to determine $\hat{\theta}$ and $\hat{\beta}$. Since the collapse data is assumed to follow a binomial distribution, the probability of k collapses out on n ground motions at the jth IM level with probability p_j is given by

$$P(k_j \text{ collapses in } n_j \text{ ground motions}) = \binom{n_j}{k_j} \cdot p_j^{k_j}(1-p_j)^{n_j-k_j} \tag{2}$$

Assuming independence in the probability of k collapses out of n ground motions, the likelihood function considering m IM levels is given by

$$L(\theta,\beta) = \prod_{j=1}^{m}\binom{n_j}{k_j}\Phi\left(\frac{\ln(IM_j/\theta)}{\beta}\right)^{k_j} \cdot \left[1-\Phi\left(\frac{\ln(IM_j/\theta)}{\beta}\right)\right]^{n_j-k_j} \tag{3}$$

Taking log on both sides of Eq. (3) gives the log-likelihood function as shown in Eq. (4). The log-likelihood function is used because it is easier to maximize using numerical optimization methods. The estimate of the fragility function parameters ($\hat{\theta}$, $\hat{\beta}$) is obtained by maximizing the log-likelihood function as shown in Eq. (5).

$$\begin{aligned}\ell(\theta,\beta) = \underset{\theta,\beta}{\arg\max}\sum_{j=1}^{m}\Bigg\{&\ln\binom{n_j}{k_j} + k_j ln\left[\Phi\left(\frac{\ln(IM_j/\theta)}{\beta}\right)\right]\\ &+(n_j-k_j)ln\left[1-\Phi\left(\frac{\ln(IM_j/\theta)}{\beta}\right)\right]\Bigg\}\end{aligned} \tag{4}$$

$$\{\hat{\theta},\hat{\beta}\} = \underset{\theta,\beta}{\arg\max}\,\ell(\theta,\beta) \tag{5}$$

To quantify parameter uncertainty, the Score, Hessian, and Information matrix are needed. For optimization and inference of Eq. (5), the Score is used. By setting the gradient equal to zero and solving for the parameters of interest, the log-likelihood function is maximized. The Hessian is a symmetric matrix of second-order partial derivatives with respect to the parameters of interest. The negative expectation of the Hessian gives the Fisher's Information matrix. Under correct probability model specification, the inverse of the Information matrix gives the covariance terms of the estimated parameters [19]. In the context of seismic collapse risk assessments, the "probability model" refers to the lognormal distribution. Even though the collapse data is characterized by the binomial distribution, this part of the likelihood function is a constant and therefore does not influence the maximization. However, as shown in Eq. (4), the likelihood maximization is affected by the lognormal distribution with feasible parameter estimates $\hat{\theta}$ and $\hat{\beta}$. In other words, the quantities of interest are the lognormal distribution parameters.

To compute the first and second-order partial derivative of the log-likelihood function, ($\ell(\theta,\beta)$), the Gauss error function [30] is needed. The log-likelihood function is re-written using the error function as shown in Eq. (6). The log-likelihood functions in Eqs. (4) and (6) are numerically equivalent.

$$\begin{aligned}\ell(\theta,\beta) = \sum_{j=1}^{m}\Bigg\{&k_j \ln\left[\frac{1}{2}\left(1+\operatorname{erf}\left(\frac{\ln\left(\frac{IM_j}{\theta}\right)}{\beta\sqrt{2}}\right)\right)\right]\\ &+(n_j-k_j)\ln\left[1-\frac{1}{2}\left(1+\operatorname{erf}\left(\frac{\ln\left(\frac{IM_j}{\theta}\right)}{\beta\sqrt{2}}\right)\right)\right]\Bigg\}\end{aligned} \tag{6}$$

In Eq. (6), $ln\binom{n_j}{k_j}$ has been neglected because it is a constant term that "drops out" when taking first and second-order partial derivatives. The Score, $S(\theta,\beta)^{\top}$, which is computed by taking the first-order partial derivative of Eq. (6), is shown in Appendix A.1. The complete derivative terms with respect to each parameter can also be found in Appendix A.1. Similarly, the Hessian, $H(\theta,\beta)$, is a square matrix obtained by taking the second-order derivative of the log-likelihood function. The matrix formulation and complete second-order derivative terms are shown in Appendix A.2. Both the Score and Hessian will later be used to calculate the sandwich estimator. As discussed earlier, Fisher's Information matrix is given by negative Hessian under correct specification i.e., $I(\theta,\beta) = -\mathbb{E}[H(\theta,\beta)]$. Subsequently, the covariance matrix of the parameter estimates is given by the inverse Information matrix as shown in Eq. (7) [19].

$$\operatorname{Var}(\theta,\beta) = I^{-1} = \begin{bmatrix}\sigma_{\theta}^2 & \sigma_{\theta,\beta}\\ \sigma_{\beta,\theta} & \sigma_{\beta}^2\end{bmatrix} \tag{7}$$

3.2. Generalized linear models

A Generalized Linear Model (GLM) with a link function is presented as an alternative to the maximum likelihood method to fitting the fragility function [3]. The inverse Probit link function links the conditional mean probability of collapse to a linear predictor (π). The inverse Probit uses the standard normal CDF which makes it equivalent to the methodology outlined in Section 3.1. Other link functions such as Logit link can also be implemented. The GLM model is presented as an easy-to-implement alternative to maximizing the log-likelihood function with lognormal CDF. The GLM estimator is readily available in popular programming languages such as Python, R, and MATLAB. It fits the fragility function by finding the coefficients of the linear predictor that maximizes the likelihood function. Most programming languages use the iteratively re-weighted least squares (IRLS) [31] optimization algorithm to find the coefficients. IRLS is essentially an L1-norm-based minimization that reduces the contribution of large residuals and improves the data fitting process. Additionally, most programs provide an option to extract the Score and Hessian which makes it an attractive alternative to MLE. Here, "MLE" is loosely used to represent the methodology described in Section 3.1. It is acknowledged that GLM also uses a maximization (or minimization of negative likelihood) algorithm and could technically be called MLE. The upside of the GLM is that it does not require an initial guess unlike MLE and the downside is that the parameters in the linear predictors (β_0, β_1) as shown in Eq. (8) are not as directly descriptive as MLE parameters, θ and β, where θ defines the IM level with 50% probability of collapse and β describes the dispersion around the median. The GLM model is defined as shown in Eq. (8)

$$\begin{aligned}P(C|IM) &= g^{-1}(\pi) = \Phi(\beta_0 + \beta_1\ln(IM)) = g^{-1}(\beta_0+\beta_1\ln(IM))\\ &= \Phi\left(\frac{\ln(IM)}{\beta} - \frac{ln(\theta)}{\beta}\right)\end{aligned} \tag{8}$$

where, $g(\cdot)$ is the link function that connects $P(C|IM)$ to the linear predictor, π. Using the lognormal CDF formulation of the probability of collapse, the linear predictor parameters, β_0 and β_1 can be transformed into lognormal CDF parameters. The transformation is performed by equating the terms inside the normal CDF, $\Phi(\cdot)$. On the left-hand side in

Eq. (8), the term with the random variable, $\beta_1 \ln(IM)$, is equated with $\ln(IM)/\beta$ on the right-hand side. Similarly, equating the intercepts on both sides yields the result as shown in Eqs. (9).

$$\theta = e^{\frac{-\beta_0}{\beta_1}} \tag{9a}$$

$$\beta = \frac{1}{\beta_1} \tag{9b}$$

One of the easy-to-use features of GLM fit is that the Hessian can easily be extracted. To quantify uncertainty in the linear predictor parameters (β_0, β_1), the inverse of the Fisher's information matrix is used. Since the primary focus of this study is to make a comparison between MLE and QMLE, the uncertainty in the linear predictor is transformed into the uncertainty in the median (θ) and standard deviation (β) using first-order Taylor approximation. In addition to transforming the uncertainty from GLM-based parameters to the MLE-based parameters, Eqs. (10a) and (10b) are also used to verify the covariance matrix obtained using the error function.

$$Var[\theta] = \sigma^2_{\beta_0}\left(\frac{\partial\theta}{\partial\beta_0}\right)^2 + \sigma^2_{\beta_1}\left(\frac{\partial\theta}{\partial\beta_1}\right)^2 + 2\sigma_{\beta_0,\beta_1}\frac{\partial\theta}{\partial\beta_0}\frac{\partial\theta}{\partial\beta_1} \tag{10a}$$

$$Var[\beta] = \sigma^2_{\beta_1}\left(\frac{\partial\beta}{\partial\beta_1}\right)^2 \tag{10b}$$

3.3. Uncertainty in collapse risk

The mean annual frequency of collapse (λ_c) is used as a risk-based performance metric. It is calculated either by integrating the continuous fragility function with the ground motion hazard curve or using the Riemann Sum as shown in Eq. (11). The Riemann sum is widely used to approximate an integral using a finite sum. To minimize the approximation error, the collapse probability is taken as the mid-point of im_i and im_{i+1} as shown in Eq. (11) below

$$\lambda_c = \int_{im} P(C|IM = im)\cdot|d\lambda_{IM}(im)| = \sum_{i=1}^{N} P(C|im)\cdot|\lambda(im_i) - \lambda(im_{i+1})| \tag{11}$$

where im is the mid-point between im_i and im_{i+1}, N is the total number of increments from IM_{min} to IM_{max}, and $\lambda_{IM}(im)$ is the mean annual frequency of exceedance of ground motion intensity, *im*. Finally, the variance of the annual collapse rate is estimated as shown in Eq. (12) [3].

$$\sigma^2_{\lambda_c} = \sum_{i=1}^{N}\sum_{j=1}^{N} |\lambda(im_i) - \lambda(im_{i+1})||\lambda(im_j) - \lambda(im_{j+1})|\rho_{ij}\sigma_{P(C|im)_i}\sigma_{P(C|im)_j} \tag{12}$$

where the variance of the probability of collapse is approximated using the first-order Taylor approximation and $\rho_{ij} = 1$.

$$\sigma^2_{P(C|im_i)} = \frac{\partial\Phi(\pi)}{\partial\pi}\sigma^2_\pi(im) \tag{13}$$

To approximate the variance of the collapse probability, the variance of the linear predictor is required. Eq. (14) [3] gives the variance of the linear predictor. It is important to note that the covariance in Eq. (7) assumes correct model specification. Eq. (14), thus, highlights the propagation of parameter estimation uncertainty into collapse risk.

$$\sigma^2_\pi(im) = \sigma^2_{\beta_0} + \sigma^2_{\beta_1}\ln(im)^2 + 2\ln(im)\sigma_{\beta_0\beta_1} \tag{14}$$

In a probabilistic sense, the collapse of a building is an event that occurs continuously and independently with a known mean frequency of occurrence (λ_c). In other words, the probability distribution of the time between two events follows a Poisson process. Since it follows a Poisson process, the probability of collapse over T years is based on the cumulative distribution function of the exponential distribution.

4. Effect of model misspecification

MLE is a flexible tool to estimate the parameters, but the fundamentals are often rooted in assumptions. The likelihood function is based on an underlying assumption that the data is known to lie within a specified parametric family of probability distributions. Misspecification can oftentimes lead to misleading inferences. In the presence of misspecification, standard tests such as the Wald test, Likelihood Ratio test, and Lagrange Multiplier/Score test are invalid. However, if a conditional mean component is correctly specified, QMLE converges to a well-defined limit and is often consistent for the parameters of interest. The variance around the parameter estimates does not equal the inverse of Fisher's information matrix. The "robust" variance estimator is computed using sandwich estimators. In this section, some of the important properties of QMLE are highlighted. These properties are the basis of MLE and are important to the validity of QMLE. The theoretical results presented in this subsection are based on White [25].

4.1. Existence of maximum likelihood parameters

Assuming that the independent random vector U has a distribution function G on Ω and a measurable Euclidean space with density g, we let f be the density drawn from a family of distribution functions $F(u,\theta)$ for every θ. Since G is unknown, F may or may not contain the true probability distribution. The quasi-log-likelihood of the sample is

$$L(U,\theta) = \sum_{i=1}^{n}\log f(U_i,\theta) \tag{15}$$

Maximizing Eq. (15) gives QMLE estimators as parameter estimates. Proof that for all n, there exists a measurable estimator, $\hat{\theta}_n$ can be found in White [25]. θ is a natural estimator for θ^*, the parameter estimate that minimizes the Kullback–Leibler Information Criterion ($I()$) also referred to as KL divergence. KL divergence characterizes how far the assumed distribution (f) is from the true distribution (g). In other words, it measures our ignorance about the true distribution. It is important to note that divergence is not the same as distance. Divergence does not meet the symmetric property of distance. Since the divergence is formulated in log space as shown in Eq. (16), the order of f and g matters. The divergence of f from g would not be the same as the divergence of g from f.

$$I(g:f,\theta) = \mathbb{E}[\log(g(U)/f(U,\theta))] = \int \log(g(u))dG(u) - \log(f(u,\theta))dG(u) \tag{16}$$

Assuming that $I(g:f,\theta)$ has a unique minimum as θ^*, the QMLE estimate, $\hat{\theta}$ converges to θ^* as $n\to\infty$. This convergence signifies that QMLE is a strongly consistent estimator for θ^*. In other words, it ensures we minimize our ignorance about the structure of the true probability model. If the probability model is correctly specified (i.e. $g(u,\theta) = f(u,\theta)$), for some θ^0, then $I(g:f,\theta)$ achieves its unique minimum at $\theta^* = \theta^0$, so that $\hat{\theta}$ is consistent for the true parameter estimates, θ^0.

4.2. The information matrix test for misspecification

White [25] showed that the QMLE is asymptotically normally distributed

$$\sqrt{n}(\hat{\theta} - \theta^*) \sim N(0, A(\theta)^{-1}B(\theta)A(\theta)^{-1}) \tag{17}$$

where $A(\theta)$ and $B(\theta)$ are called sandwich estimators and can be computed as

$$A(\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 log f(u_t,\theta)}{\partial\theta_i\partial\theta_j} = H(\theta) \tag{18a}$$

$$B(\theta) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial log f(u_t,\theta)}{\partial\theta_i}\cdot\frac{\partial log f(u_t,\theta)}{\partial\theta_j} = S(\theta)S(\theta)^\top \tag{18b}$$

Table 1
Description of wood-frame buildings.

Building name	No. of stories	Description
B1-Existing	1	Stucco exterior, gypsum wallboard interior with 2 ft high stucco cripple wall
B1-Retrofitted	1	B1-Existing retrofitted with 2 ft tall wood-frame shear wall
B2-Existing	1	Stucco exterior, gypsum wallboard interior with 6 ft high stucco cripple wall
B2-Retrofitted	1	B2-Existing retrofitted with 6 ft tall wood-frame shear wall
B3-Existing	2	Stucco exterior, gypsum wallboard interior with 2 ft high stucco cripple wall
B3-Retrofitted	2	B3-Existing retrofitted with 2 ft tall wood-frame shear wall
B4-Existing	2	Stucco exterior, gypsum wallboard interior with 6 ft high stucco cripple wall
B4-Retrofitted	2	B4-Existing retrofitted with 6 ft tall wood-frame shear wall

The information matrix equivalence theorem essentially says that when the model is correctly specified, the information matrix can be expressed in either Hessian form, $-A(\theta)$ or the outer product form $B(\theta)$. In other words, the equality condition to identify model misspecification can be formulated simply as

$$A(\theta) + B(\theta) = 0 \tag{19}$$

When the equality holds, the asymptotic normality shown in Eq. (17) simplifies to

$$\sqrt{n}(\hat{\theta} - \theta^{*}) \sim N(0, A(\theta)^{-1}) \tag{20}$$

When the equality in Eq. (19) fails, it is concluded that the model is misspecified. One of the many ways to check the severity of misspecification is to analyze the variance estimator obtained from $A(\theta)^{-1}B(\theta)A(\theta)^{-1}$.

5. Case study

To study the effect of model misspecification, collapse data from MSA of eight wood-frame buildings are used. The single-family archetypes were developed as part of the Pacific Earthquake Engineering Research (PEER) Center and California Earthquake Authority (CEA) project [32] on cripple wall residential buildings. The objective of the project was to quantify the benefit of retrofitting unbraced cripple walls in high seismic regions. The overall building dimension was 40 ft × 30 ft. As part of the PEER-CEA project, various structural properties that influence the performance of the building such as cripple wall height, interior panel sheathing material, seismic weight, and the number of stories were considered to create 64 distinct building archetypes [33]. Table 1 summarizes the characteristics of the four variants considered in the current study. Wood structural panels are used to retrofit the unbraced and unbolted cripple wall, which represents a soft-story vulnerability. For the retrofitted buildings, the wood structural panels are placed at the two ends of each line of cripple walls. The length of the structural panels varies based on the FEMA P-1100 specification [34]. Additional details about the variants and their associated retrofits can be found in Reis [33] and Welch and Deierlein [32].

Three-dimensional structural models of one-story and two-story cripple wall buildings are constructed and nonlinear dynamic analyses are performed using the Open System of Earthquake Engineering Simulation (*OpenSEES*) [35]. Hysteretic shear spring elements are placed at the centroid of the cripple walls to simulate their force–displacement behavior. Two parallel Pinching4 hysteretic springs are used to capture the wide range of cyclic behavior between the near elastic (small displacement) response and the highly nonlinear post-peak behavior at large displacements. Each building was subjected to 16 different hazard levels with 45 ground motions at each individual hazard level. The spectral acceleration corresponding to a period of 0.25 s ($Sa_{0.25s}$) is used as the IM. The ground motions were also selected as a part of the PEER-CEA project for a set of ten sites [36]. The ground motion records used to produce analysis results are based on a San Francisco site with stiff soil (Soil D) and a target upper 30-meter shear wave velocity (V_{S30}) of 270 m/s. Fig. 1 summarizes the median collapse intensity and dispersion values obtained by implementing MLE. As expected, the retrofitted buildings have higher median collapse capacities. The dispersion, which quantifies record-to-record uncertainty, is generally comparable between the existing and retrofitted cases. It is interesting to note that the median collapse capacity of the retrofitted two-story building with 6 ft tall wood-frame shear wall (B4) is almost the same as the retrofitted two-story building with 2 ft tall shear wall (B3).

Assuming correct specification, variance estimators are computed using Eq. (7). The variance around the parameter estimates (i.e., $\mathrm{Var}(\theta)$ and $\mathrm{Var}(\beta)$) is summarized in Fig. 2. For comparison, a standardized measure of variability, coefficient of variation (CV) is used. The CV is calculated by taking the ratio of standard deviation to the mean values. Comparable CV of the median collapse intensity for the existing and the retrofitted buildings as shown in Fig. 2(a) implies that retrofitting the building neither increases or decreases the relative dispersion around the median value. With the exception of building B2, Fig. 2(b) shows that a similar conclusion can be drawn about the relative dispersion around the log standard deviation.

To better understand the uncertainty around the parameter estimates, data from building B2 is used to plot their probability density function (PDF). Figs. 3(a) and 3(b) show the PDF of the median and dispersion respectively. The result shows that most of the probability density is concentrated around the mean value in the existing case. For the retrofitted case, the probability density is more distributed implying higher levels of dispersion, though the increase is not substantive. Similarly, the PDF of the log standard deviation shows a slightly broadened tail for the retrofitted case which explains the higher CV as highlighted in Fig. 2(b). Analogous results were obtained for the rest of the buildings as well where slightly spiked probability density was observed for the median value and more evenly distributed probability density for the log standard deviation. Fig. 3 also shows the PDF (dashed lines) of the parameter estimates obtained using the sandwich estimators.

5.1. Effect of model misspecification in fragility function fitting

As shown in Eq. (5), MLE is implemented to estimate the lognormal distribution parameters $(\hat{\theta}, \hat{\beta})$. Following the setup outlined in Appendix B, the score and the Hessian are computed. Assuming that the lognormal probability distribution is the "true" probability model (i.e., correct specification), an asymptotic covariance matrix is computed. As shown in Eq. (20), the covariance matrix equals the inverse of the negative Fisher's Information matrix. With known variability in the parameter estimates, the uncertainty in the fragility curve is quantified by establishing a confidence interval. Fig. 4. shows the fitted fragility curve in black for building B2-Existing along with the empirical probability of collapse data points in green. It is found that the 95% confidence interval provides a good range by bounding almost all of the empirical data.

In addition to quantifying the parameter estimation uncertainty, several other techniques are also presented as an alternative to computing the Hessian matrix. The second-order derivative terms of the error function-based likelihood with respect to its parameters are used to formulate the Hessian matrix. The purpose of presenting multiple methods of computing the Score and the Hessian is two-fold: (1) verifying the rigorousness of the error function formulation and (2) presenting an alternative methodology that can be easily implemented in programming languages such as Python, R, and MATLAB. In addition to

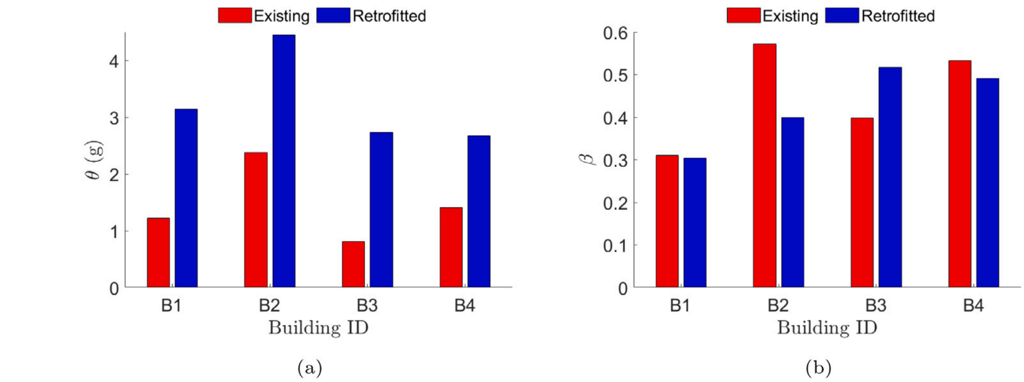


Fig. 1. (a) Median and (b) log standard deviation of the collapse intensity for all four existing and retrofitted building cases.

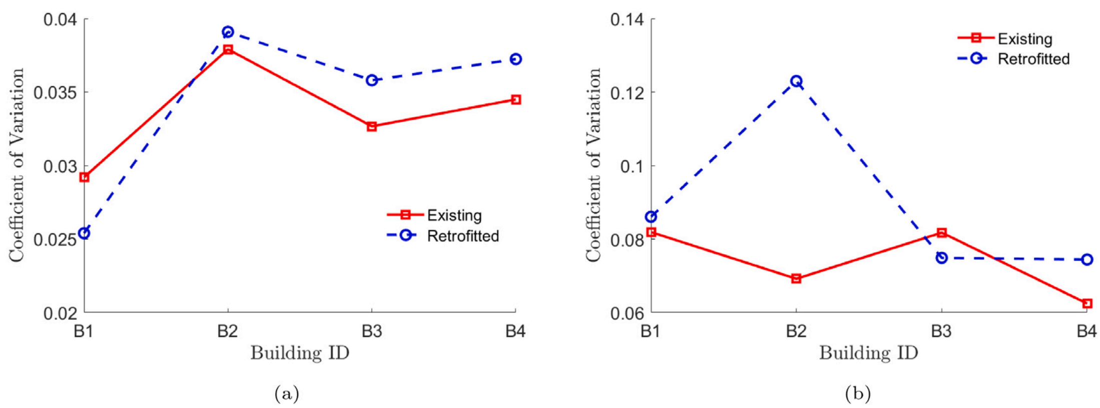


Fig. 2. Coefficient of variation for (a) the median and (b) log standard deviation under correct specification assumption.

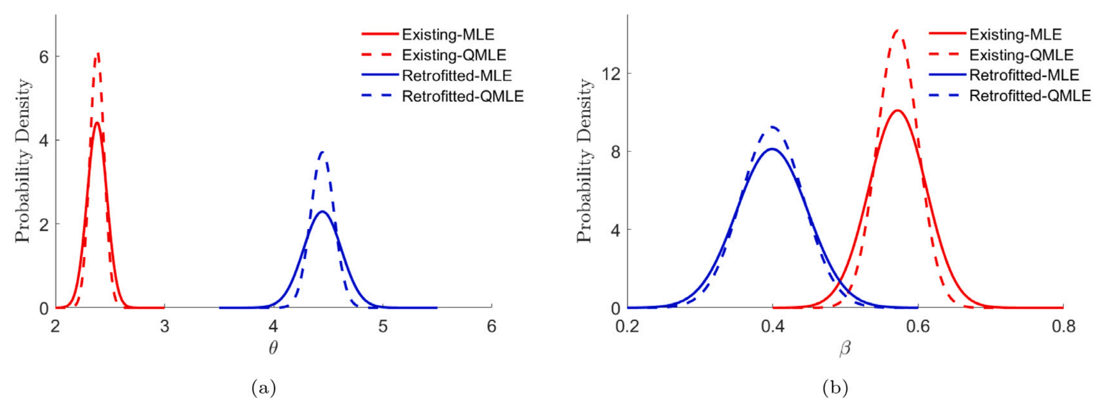


Fig. 3. Normal PDF showing the distribution of the (a) median and (b) log standard deviation of the collapse capacity for the B2 building.

computing the error function-based covariance matrix, two supplemental methodologies to compute covariances are also presented: (1) Taylor approximation (2) Numerical integration using built-in packages.

The Taylor series expansion is used to obtain a first-order second-moment approximation of the parameter estimates. Eqs. (10)–(14) shows the steps used to estimate the variance in θ and β. The values obtained should approximately be equal to the variance terms as shown in Appendix B.2. Additionally, built-in numerical differentiation tools in programming languages such as Python and R are also used to extract the Hessian. In Python, the "numdifftools" package [37] is used to estimate the Hessian and the Score (Jacobian). Numdifftools uses the complex-step method of numerical differentiation. Similarly, in R, the optimization package "optim" [38] returns a numerically differentiated Hessian matrix. For all eight buildings analyzed, the Hessian matrix is found to be consistent regardless of the method, or the built-in package used to compute it.

After the Hessian and the Score are verified to be correct, they are used to assess the equality condition outlined in Eq. (19). The sandwich estimators shown in Appendix B.1 demonstrate the result obtained for B2-Existing. The results show that the equality condition does not hold, i.e. $A(\theta) + B(\theta) \neq 0$. It is found that none of the eight buildings meet the equality condition. The result signifies the prevalence of model misspecification in the maximum likelihood method used to get the lognormal distribution parameters (θ, β).

The severity of misspecification is determined by analyzing the values obtained from $A(\theta)^{-1}B(\theta)A(\theta)^{-1}$. Appendix B.2 shows the covariance matrix computed based on the inverse of the Fisher's information matrix (MLE) and the sandwich estimators (QMLE). The comparison shows that the parameter variances obtained using sandwich estimators

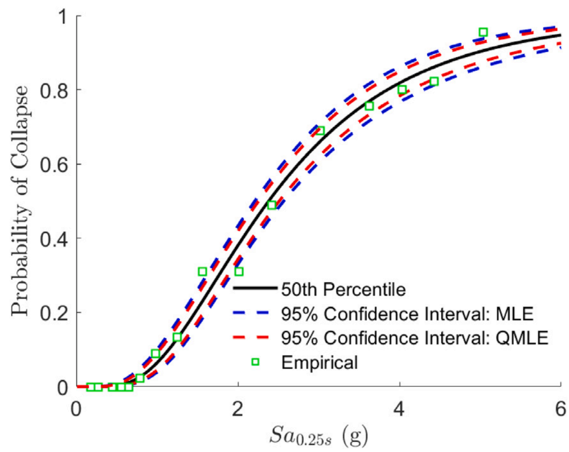


Fig. 4. Fitted fragility curve for an example building (B2-existing) with 95% confidence interval.

are approximately half that of the case where the inverse Fisher's information matrix is used. Fig. 5 summarizes the CV values for the respective parameter estimates using the sandwich estimators. On average, the CV of the median for the existing building tends to be higher compared to the retrofitted counterpart while the CV of the logarithmic standard deviation does not follow any specific pattern. The results from Fig. 2(a) and Fig. 5(a) show that the CV of the median collapse capacity obtained from the sandwich estimator (QMLE) ranges from 42% to 70% that of the MLE case. Similarly, Figs. 2(b) and 5(b) indicate that the CV of the log standard deviation computed using QMLE is 48–88% of the corresponding values from MLE. In other words, the parameter estimation dispersion in $\hat{\theta}$ and $\hat{\beta}$ is generally lower when probability model misspecification is considered.

5.2. Effect of model misspecification in collapse risk

As shown in Eq. (11), the mean annual frequency of collapse is computed using the estimated fragility function parameters. Fig. 6 shows the estimated value of the collapse rate for all buildings (existing and retrofitted). Consistent with our prior knowledge of the physical implications, the retrofitted buildings have lower collapse rates which indicate enhanced performance. We also propagate the parameter estimation uncertainty in the fragility to the collapse rate. More specifically, the covariance matrix corresponding to θ and β obtained from the inverse Fisher's information matrix is used to quantify the parameter estimation uncertainty in the collapse rate (Eqs. (12)–(14)). To assess the impact of model misspecification, the covariance matrix obtained from the sandwich estimator is also used. Fig. 7 summarizes the CV values for the collapse rate from MLE and QMLE. Fig. 7(a) shows the results computed assuming the correct specification (MLE) while Fig. 7(b) summarizes the results obtained using the sandwich estimator (QMLE). Both figures show that the standardized dispersion in the collapse rate appears to be higher in all the retrofitted cases as compared to the existing buildings. In terms of the effect of misspecification, the net result is that the CV in the collapse rate obtained from QMLE ranges from approximately 40–80% of the MLE values.

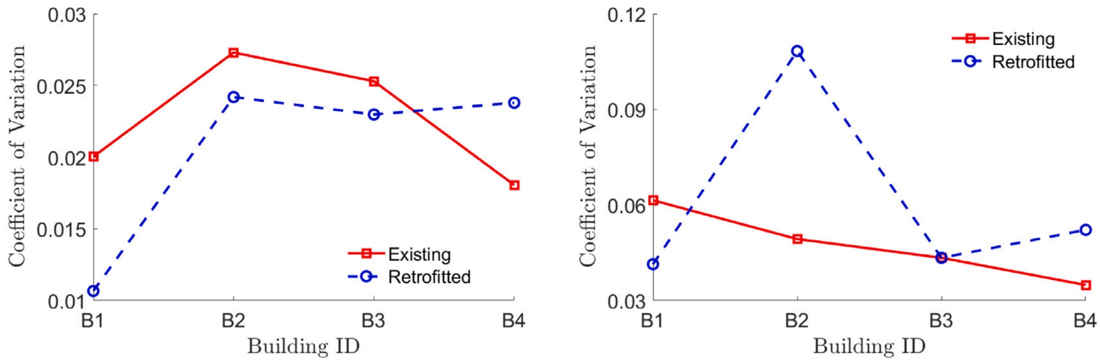
5.3. Bootstrapping

In addition to the procedures described in Section 3, sampling or resampling techniques are used to quantify the uncertainty in the collapse risk. Monte Carlo simulation is used to draw random samples utilizing the fragility parameter and covariance estimates as shown in Eqs. (9) and (10). The Monte Carlo method can be used as a stand-alone technique. However, in this study, the simulated samples are resampled with replacement. In essence, the random sample generated from Monte Carlo is treated as a population from which a finite number of samples are drawn. It is important to note that the variance is inversely related to the number of realizations. In other words, it is possible to fine-tune the variability in the random samples by drawing a large number of realizations. To negate subjectivity, 500 samples are drawn as was done in previous studies [6,20]. Fig. 8(a) shows the empirical distribution of fragility curves obtained from Monte Carlo simulation for the B2-Existing building and Fig. 8(b) presents a comparison for the standard-deviation of the collapse rate for all buildings. In Fig. 8(a), we see that central tendency of the Monte Carlo samples average to the empirical fragility curve shown in Fig. 4. Also, Fig. 8(b) highlights that the standard deviation computed using Eq. (12) is generally equivalent to the one obtained from the bootstrapping method.

6. Conclusion

This paper discussed the effect of probability model misspecification on fragility function parameter estimates (see Table 2). More specifically, the effect of model misspecification is propagated into seismic collapse performance and risk assessment to study the effect on parameter estimation uncertainty. The successful application of quasi-maximum likelihood estimation (QMLE) requires "sandwich" estimators that rely on the Jacobian (Score) and Hessian of the likelihood function. The Hessian is computed using the error function and is verified using two different approaches (1) Taylor approximation (2) Numerical integration using a built-in software package. The information matrix equality test performed on collapse data from single family woodframe buildings showed that model misspecification is present. The implication to the collapse fragility was evaluated by comparing the coefficient of variation (CV) of the median ($\hat{\theta}$) and log standard deviation ($\hat{\beta}$), which represent the parameter estimation uncertainty. For $\hat{\theta}$, the CV obtained from QMLE, was, on average, 38% lower than the case where maximum likelihood estimation (MLE) is used. Similarly, the average ratio between the CV of $\hat{\beta}$ from QMLE and MLE was 64%. The fragility parameter estimation uncertainty was also propagated into the collapse risk assessment using analytical (Riemann Sum) and sampling (Monte Carlo Simulation) approaches, which produced comparable results. Compared to when MLE was used, the CV of the mean annual frequency of collapse due to parameter estimation uncertainty was computed to be 17–58% lower in the QMLE case.

MLE and other parameter estimation approaches such as Generalized Linear Models are commonly used in research and practice with the assumption that the probability model is correctly specified which might not always be the case. Oftentimes, the assumption about the "true" model introduces additional uncertainty in the parameter estimates. This paper demonstrated the use of the Information equality test to identify probability model misspecification and quantify its effect on collapse risk. In the current study, a robust consideration of misspecification was found to generally reduce the parameter estimation uncertainty in collapse risk compared to when the correct probability model is assumed. However, it is important to note that only a single construction type (woodframe residential buildings) and limit state (collapse) was considered. Additional studies are needed where the effect of probability model misspecification on parameter estimation uncertainty for other construction types (e.g., steel and concrete moment frames) and limit states (e.g., demolition, post-earthquake safety and functionality) is evaluated. If there are situations where a robust consideration of model misspecification leads to an increase in parameter estimation dispersion (e.g., [39]), this additional uncertainty should be considered in the associated risk assessment.



(a) (b)

Fig. 5. Coefficient of variation of the (a) median and (b) log standard deviation computed using sandwich estimators.

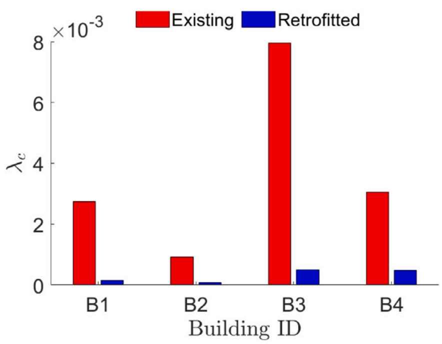


Fig. 6. Estimated mean annual frequency of collapse for all building cases.

Table 2
Summary of important theorems implemented in QMLE.

Theorem	Description
Consistency Theorem	Assuming that (1) the probability density, $f(U_i, \theta)$ is observable and (2) the density is twice continuously differentiable on θ, as $n \to \infty$, $\hat{\theta}_n \to \theta^*$. The QMLE is a consistent estimator of the "true" parameter values (θ^*) which minimizes the KL divergence.
Asymptotic Distribution Theorem	As $n \to \infty$, $\sqrt{n}(\hat{\theta}_n - \theta^*)$ converges to a zero-mean Gaussian distribution with non-singular covariance matrix $A(\theta)^{-1}B(\theta)A(\theta)^{-1}$. If the probability model is correctly specified i.e., $A(\theta) = -B(\theta)$, the covariance matrix is simply represented by $A(\theta)^{-1}$. Hence, the asymptotic distribution is a zero-mean Gaussian with $A(\theta)^{-1}$ covariance matrix.
The Information Matrix Equivalence Theorem	If the Hessian matrix-based covariance matrix ($A^{-1}(\theta)$) and gradient-based covariance matrix ($B^{-1}(\theta)$) are asymptotically different, it indicates the presence of model misspecification. In other words, $A(\theta) + B(\theta) = 0$ can be used as a basis for the identification of the probability model misspecification.

Software tools

For those interested in adopting the methodology, codes used to produce results in this paper are available at GitHubPage. The code utilizes an object-oriented programming paradigm to make the application of collapse risk assessment seamless. The code implements MLE and GLM in addition to providing an option to conduct bootstrapping.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to acknowledge the University of California Office of the President, United States for generous UC-HBCU Fellowship offered to the first author.

Appendix A. Maximum likelihood equations using the error function

$$\text{Let, } E = e^{-\frac{\ln\left(\frac{IM_j}{\theta}\right)^2}{2\beta^2}} \quad ; \quad Q = e^{-\frac{\ln\left(\frac{IM_j}{\theta}\right)^2}{\beta^2}} \quad ; \quad P = \operatorname{erf}\left(\frac{\ln\left(\frac{IM_j}{\theta}\right)}{\sqrt{2}\beta}\right)$$

A.1. Score

$$\frac{\partial \ell(\theta,\beta)}{\partial \theta} = \frac{\sqrt{2}}{\sqrt{\pi}\theta\beta}\sum_{j=1}^{m} -k_j\frac{E}{1+P} + (n_j - k_j)\frac{E}{1-P}$$

$$\frac{\partial \ell(\theta,\beta)}{\partial \beta} = \frac{\sqrt{2}}{\sqrt{\pi}\beta^2}\sum_{j=1}^{m} -k_j\frac{\ln\left(\frac{IM_j}{\theta}\right)E}{1+P} + (n_j - k_j)\frac{\ln\left(\frac{IM_j}{\theta}\right)E}{1-P}$$

$$S(\theta,\beta)^{\top} = \left[\frac{\partial \ell(\theta,\beta)}{\partial \theta}, \frac{\partial \ell(\theta,\beta)}{\partial \beta}\right]$$

A.2. Hessian

$$\begin{aligned}\frac{\partial^2 \ell(\theta,\beta)}{\partial \theta^2} = \frac{1}{\theta^2\beta}\sum_{j=1}^{m} & k_j\frac{\sqrt{2}E}{\sqrt{\pi}(P+1)} - k_j\frac{2Q}{\beta\pi(P+1)^2} - k_j\frac{\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)E}{\beta^2\sqrt{\pi}(P+1)} \\ & - (k_j - n_j)\frac{\sqrt{2}E}{\sqrt{\pi}(P-1)} + (k_j - n_j)\frac{2Q}{\beta\pi(P-1)^2} \\ & + (k_j - n_j)\frac{\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)E}{\beta^2\sqrt{\pi}(P-1)}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell(\theta,\beta)}{\partial \theta^2} = \frac{1}{\beta^3}\sum_{j=1}^{m} & k_j\frac{2\sqrt{2}E}{\sqrt{\pi}(P+1)} - k_j\frac{2\ln\left(\frac{IM_j}{\theta}\right)^2 Q}{\beta\pi(P+1)^2} - k_j\frac{\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)^3 E}{\beta^2(P+1)} \\ & - (k_j - n_j)\frac{2\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)E}{\sqrt{\pi}(P-1)} + (k_j - n_j)\frac{2\ln\left(\frac{IM_j}{\theta}\right)^2 Q}{\beta\pi(P-1)^2} \\ & + (k_j - n_j)\frac{\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)^3 E}{\sqrt{\pi}\beta^2(P-1)}\end{aligned}$$

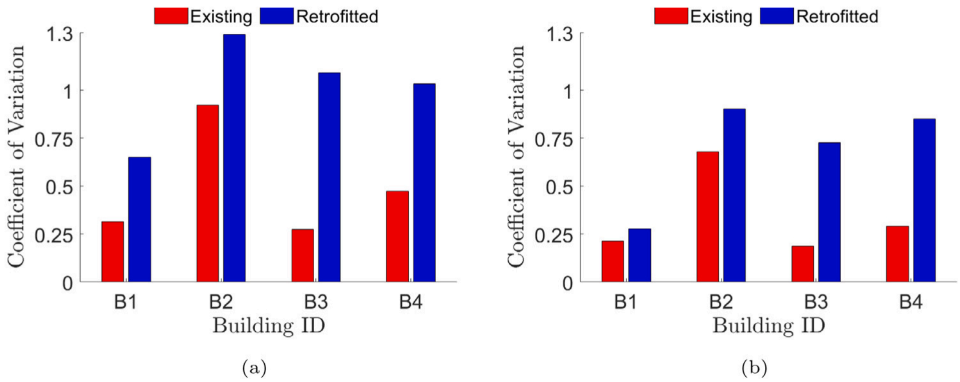


Fig. 7. Comparison of coefficient of variation of the collapse rate computed using (a) asymptotic variance estimators and (b) sandwich estimators.

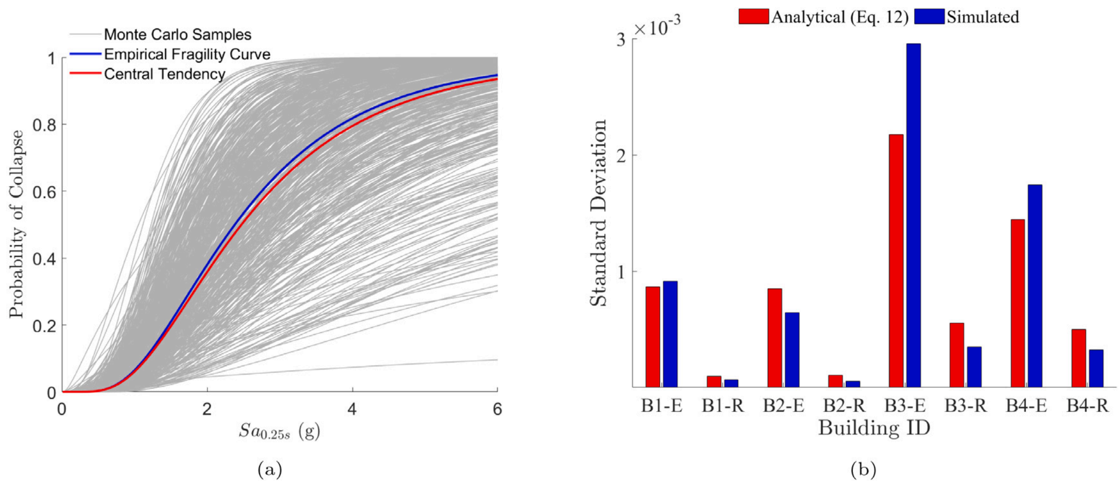


Fig. 8. (a) Monte Carlo fragility samples for B2-Existing. (b) Comparison of standard-deviation computed using Taylor approximation and bootstrapping method.

$$
\begin{aligned}
\frac{\partial^2 \ell(\theta,\beta)}{\partial\theta\partial\beta} = \frac{1}{\theta\beta^2}\sum_{j=1}^{m} & k_j\frac{\sqrt{2}E}{\sqrt{\pi}(P+1)} - k_j\frac{2\ln\left(\frac{IM_j}{\theta}\right)Q}{\beta\pi(P+1)^2} - k_j\frac{\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)^2 E}{\beta^2\sqrt{\pi}(P+1)} \\
& -(k_j-n_j)\frac{\sqrt{2}E}{\sqrt{\pi}(P-1)} + (k_j-n_j)\frac{2\ln\left(\frac{IM_j}{\theta}\right)Q}{\beta\pi(P-1)^2} \\
& +(k_j-n_j)\frac{\sqrt{2}\ln\left(\frac{IM_j}{\theta}\right)^2 E}{\beta^2\sqrt{\pi}(P-1)}
\end{aligned}
$$

$$
H(\theta,\beta) = \begin{bmatrix} \frac{\partial^2\ell(\theta,\beta)}{\partial\theta^2} & \frac{\partial^2\ell(\theta,\beta)}{\partial\theta\partial\beta} \\ \frac{\partial^2\ell(\theta,\beta)}{\partial\beta\partial\theta} & \frac{\partial^2\ell(\theta,\beta)}{\partial\beta^2} \end{bmatrix}
$$

$$
\frac{\partial\theta}{\partial\beta_0} = \frac{-e^{\frac{-\beta_0}{\beta_1}}}{\beta_1} \qquad \frac{\partial\theta}{\partial\beta_1} = \frac{\beta_0 e^{\frac{-\beta_0}{\beta_1}}}{\beta_1^2}
$$

$$
\frac{\partial\beta}{\partial\beta_1} = -\frac{1}{\beta_1^2}
$$

Appendix B. An example implementation of the equality test

MLE and QMLE is implemented using the collapse data of the building B2-Existing.

B.1. Sandwich estimators

The sandwich estimators ($A(\theta)$ and $B(\theta)$) are computed using Eq. (18)

$$
A(\theta) = \begin{bmatrix} 122.7572 & -5.5052 \\ -5.5052 & 638.9830 \end{bmatrix} \qquad B(\theta) = \begin{bmatrix} 63.6039 & -14.1952 \\ -14.1952 & 324.0924 \end{bmatrix}
$$

B.2. Variance estimators

The following matrices represent the variance estimators. $A(\theta)^{-1}$ is the variance estimator under correct specification (MLE). $A(\theta)^{-1}B(\theta)A(\theta)^{-1}$ is the variance estimator for QMLE.

$$
A(\theta)^{-1} = \begin{bmatrix} 8.1493e^{-03} & 7.0211e^{-05} \\ 7.0211e^{-05} & 1.5655e^{-03} \end{bmatrix}
$$

$$
A(\theta)^{-1}B(\theta)A(\theta)^{-1} = \begin{bmatrix} 4.2240e^{-03} & -6.9977e^{-08} \\ -6.9977e^{-08} & 7.9438e^{-04} \end{bmatrix}
$$

References

[1] FEMA. Seismic performance assessment of buildings, vol. 1 - Methodology FEMA P-58 Second Edition. Washington, D.C.: Federal Emergency Management Agency; 2009.

[2] Baker JW. Efficient analytical fragility function fitting using dynamic structural analysis. Earthq Spectra 2015;31(1):579–99.

[3] Lallemant D, Kiremidjian A, Burton H. Statistical procedures for developing earthquake damage fragility curves. Earthq Eng Struct Dyn 2015;44(9):1373–89.

[4] Porter K, Kennedy R, Bachman R. Creating fragility functions for performance-based earthquake engineering. Earthq Spectra 2007;23(2):471–89.

[5] Shinozuka M, Feng MQ, Lee J, Naganuma T. Statistical analysis of fragility curves. J Eng Mech 2000;126(12):1224–31.

[6] Trevlopoulos K, Zentner I. Seismic fragility curve assessment based on synthetic ground motions with conditional spectra. Pure Appl Geophys 2019;1–16.
[7] Bradley BA, Dhakal RP. Error estimation of closed-form solution for annual rate of structural collapse. Earthq Eng Struct Dyn 2008;37(15):1721–37.
[8] Vamvatsikos D, Cornell CA. Applied incremental dynamic analysis. Earthq Spectra 2004;20(2):523–53.
[9] Jalayer F. Direct probabilistic seismic anaysis: implementing non-linear dynamic assessments [Ph.D. thesis], Stanford University Stanford, CA; 2003.
[10] FEMA. Quantification of building seismic performance factors: Component equivalency methodology, vol. FEMA P-695. Washington, D.C.: Federal Emergency Management Agency; 2011.
[11] Bradley BA. A critical examination of seismic response uncertainty analysis in earthquake engineering. Earthq Eng Struct Dyn 2013;42(11):1717–29.
[12] Haselton C, Deierlein G. Assessing seismic collapse safety of modern reinforced concrete frame buildings. PEER Rep 2007;8.
[13] Dolsek M. Incremental dynamic analysis with consideration of modeling uncertainties. Earthq Eng Struct Dyn 2009;38(6):805–25.
[14] Liel AB, Haselton CB, Deierlein GG, Baker JW. Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. Struct Saf 2009;31(2):197–211.
[15] Asgarian B, Ordoubadi B. Effects of structural uncertainties on seismic performance of steel moment resisting frames. J Construct Steel Res 2016;120:132–42.
[16] Gokkaya BU, Baker JW, Deierlein GG. Quantifying the impacts of modeling uncertainties on the seismic drift demands and collapse risk of buildings with implications on seismic design checks. Earthq Eng Struct Dyn 2016;45(10):1661–83.
[17] Yin Y-J, Li Y. Seismic collapse risk of light-frame wood construction considering aleatoric and epistemic uncertainties. Struct Saf 2010;32(4):250–61.
[18] Lallemant D, Kiremidjian A. Accounting for uncertainty in earthquake fragility curves. In: 16th world conference on earthquake engineering. 2017.
[19] Fisher RA. On the mathematical foundations of theoretical statistics. Phil Trans R Soc Lond Ser A Contain Pap Math Phys Charact 1922;222(594–604):309–68.
[20] Iervolino I. Assessing uncertainty in estimation of seismic response for PBEE. Earthq Eng Struct Dyn 2017;46(10):1711–23.
[21] Oehlert GW. A note on the delta method. Amer Statist 1992;46(1):27–9.
[22] Eads L, Miranda E, Krawinkler H, Lignos DG. An efficient method for estimating the collapse risk of structures in seismic regions. Earthq Eng Struct Dyn 2013;42(1):25–41.
[23] Pourreza F, Mousazadeh M, Basim MC. An efficient method for incorporating modeling uncertainties into collapse fragility of steel structures. Struct Saf 2021;88:102009.
[24] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22(1):79–86.
[25] White H. Maximum likelihood estimation of misspecified models. Econometrica 1982;1–25.
[26] Yi Z. Performance-based analytics-driven seismic retrofit of woodframe buildings [Ph.D. thesis], University of California, Los Angeles; 2020.
[27] Le Cam L. On the asymptotic theory of estimation and testing hypotheses. In: Proceedings of the third berkeley symposium on mathematical statistics and probability, Volume 1: Contributions to the theory of statistics. The Regents of the University of California; 1956.
[28] Wald A. Note on the consistency of the maximum likelihood estimate. Ann Math Stat 1949;20(4):595–601.
[29] Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the fifth berkeley symposium on mathematical statistics and probability, vol. 1, no. 1. University of California Press; 1967, p. 221–33.
[30] Andrews LC. Special functions of mathematics for engineers, vol. 49. Spie Press; 1998.
[31] Green PJ. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. J R Stat Soc Ser B Stat Methodol 1984;46(2):149–70.
[32] Welch DP, Deierlein GG. Technical background report for structural analysis and performance assessment. PEER Report No. 2020/22, University of California, Berkeley, CA: Pacific Earthquake Engineering Research Center; 2020.
[33] Reis E. Seismic performance of single-family wood-frame houses: Comparing analytical and industry catastrophe models. PEER Report No. 2020/24, University of California, Berkeley, CA: Pacific Earthquake Engineering Research Center; 2020.
[34] FEMA. Vulnerability-based seismic assessment and retrofit of one- and two-family dwellings. Volume 1- Prestandard, (FEMA P-1100). Washington, D.C.: Federal Emergency Management Agency; 2019.
[35] Mazzoni S, McKenna F, Scott MH, Fenves GL, et al. Opensees command language manual. Pac Earthq Eng Res (PEER) Center 2006;264.
[36] Mazzoni S, Gregor N, Al Atik L, Bozorgnia Y, Welch DP, Deierlein GG. Probabilistic seismic hazard analysis and selecting and scaling ground-motion records. PEER Report No. 2020/14, University of California, Berkeley, CA: Pacific Earthquake Engineering Research Center; 2020.
[37] Walter SF, Lehmann L. Algorithmic differentiation in Python with AlgoPy. J Comput Sci 2013;4(5):334–44.
[38] Nash JC, Varadhan R, et al. Unifying optimization algorithms to aid software system users: optimx for R. J Stat Softw 2011;43(9):1–14.
[39] Fay MP, Graubard BI, Freedman LS, Midthune DN. Conditional logistic regression with sandwich estimators: application to a meta-analysis. Biometrics 1998;195–208.