DEVELOPING RELIABLE SEISMIC DEMAND MODELS WITH LIMITED DATA

M. Kenawy¹, A. Ahmadi² and S. Kunnath³

ABSTRACT

The development of a reliable demand model requires extensive simulations, typically requiring thousands of ground motions for a given scenario. However, in an engineering context for preliminary seismic assessment using a limited set of simulations, it is necessary that the seismic demand model developed for a specific structure is generally unaffected by any bias in the calibration dataset. Two approaches for developing the demand model are investigated: regular least squares regression and penalized least squares (or Ridge regression). The effectiveness of the latter approach is demonstrated through a case study wherein seismic demand models are developed to estimate the peak interstory drift of a six-story steel moment frame. Nonlinear simulations using 36 records and three scale factors (resulting in a total of 108 analyses), selected to be compatible with the conditional mean spectrum corresponding to 2%/50 year hazard for the site, are carried out and the performance of the two modeling methods is investigated by randomly selecting several realizations of 10 sample simulations. It is found that the predictive ability and variance of a demand model using standard linear regression is unreliable. On the other hand, biased estimation of the regression coefficients using Ridge regression is shown to remedy this problem by reducing the variance in the estimated regression coefficients.

¹ Graduate Student Researcher, Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616, USA. (email: mmkenawy@ucdavis.edu)
² Former Graduate Student, Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616.
³ Professor, Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616, USA.

Developing Reliable Seismic Demand Models with Limited Data

M. Kenawy¹, A. Ahmadi² and S. Kunnath³

ABSTRACT

The development of a reliable demand model requires extensive simulations, typically requiring thousands of ground motions for a given scenario. However, in an engineering context for preliminary seismic assessment using a limited set of simulations, it is necessary that the seismic demand model developed for a specific structure is generally unaffected by any bias in the calibration dataset. Two approaches for developing the demand model are investigated: regular least squares regression and penalized least squares (or Ridge regression). The effectiveness of the latter approach is demonstrated through a case study wherein seismic demand models are developed to estimate the peak interstory drift of a six-story steel moment frame. Nonlinear simulations using 36 records and three scale factors (resulting in a total of 108 analyses), selected to be compatible with the conditional mean spectrum corresponding to 2%/50 year hazard for the site, are carried out and the performance of the two modeling methods is investigated by randomly selecting several realizations of 10 sample simulations. It is found that the predictive ability and variance of a demand model using standard linear regression is unreliable. On the other hand, biased estimation of the regression coefficients using Ridge regression is shown to remedy this problem by reducing the variance in the estimated regression coefficients.

Introduction

In Performance-Based Earthquake Engineering, structural demands are typically represented by a distribution with an appropriate central tendency (e.g., median) and a dispersion value which captures the randomness of the seismic activity. Previous studies [1, 2] have shown that a demand model which includes an adequate number of spectral accelerations (at selected modal periods) sufficiently represents the characteristics of ground motion, and can reasonably predict the primary structural response quantity of interest or engineering demand parameter (EDP). Maximum interstory drift (MIDR) is the most widely used EDP to characterize the response of the structure to the expected seismic loading and regression analysis is the commonly employed approach to generate the predictive equation that relates the ground motion properties to the structural response measure (EDP). A meaningful and stable demand model should also predict the response of the structure to other excitations which are not part of the calibration data but which belong to the same hazard scenario. Another criterion that is important in the context of a preliminary seismic assessment is

---

¹ Graduate Student Researcher, Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616, USA. (email: mmkenawy@ucdavis.edu)
² Former Graduate Student, Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616.
³ Professor, Department of Civil and Environmental Engineering, University of California, Davis, CA, 95616, USA.

the sensitivity of the model to the sample data size. This study focuses primarily on the effectiveness of demand models when dealing with limited data sets. Two approaches for developing the demand model are investigated: the commonly used least squares regression and an enhanced approach based on penalized least squares, also referred to as Ridge regression. These approaches are applied to data sets of a predicted structural response parameter (MIDR) of a six-story steel moment frame building subjected to a set of 108 ground motion records. The performance of the demand models is evaluated by comparing the predictive capability of each method for several realizations of 10 sample simulations.

**Building Details and Ground Motion Selection**

The structural model is a 2-D frame of an existing 6-story steel perimeter moment-resisting frame structure (with modal periods $T_{1,2,3} = 1.39, 0.51,$ and 0.3 sec) in Burbank California (Figure 1.a). Details of the building can be found in [3]. The building is instrumented by the California Strong Motion Instrumentation Program (CSMIP), and recorded data are available for several past earthquakes. The simulation model was calibrated to the observed response data to validate the model. The nonlinear response of the elements was modeled using distributed plasticity. The force-based element in OpenSees [4] with Modified Gauss-Radau plastic hinge integration (forceBeamColumn) was used for frame members. A leaning column was added to the model to account for the amplification of the sidesway of the structure caused by gravity loads (i.e., the gravity load carried by the interior columns that are not part of the lateral load resisting system).

![Figure 1](image)

(a) 2-D frame model of the 6-story steel moment-frame building; (b) Spectra of selected ground motions

The ground motions used in the simulations were randomly selected from a set of 36 ground motions which were "scale-matched" to a Conditional Mean Spectrum corresponding to a 2%/50 year hazard level for a site in Burbank, California. The ground motions (shown in Figure 1.b) were obtained from the PEER NGA database. The ground motion selection code available at [https://github.com/bakerjw/CS_Selection](https://github.com/bakerjw/CS_Selection) was used for the selection process. The 36 ground motions were scaled to three different levels to produce a range of linear/nonlinear responses resulting in a total of 108 simulations.
Demand Modeling: Least Squares versus Ridge Regression

One of main issues with regular regression using least squares is that the identified regression coefficients can vary significantly from one sample set to the next, even for the same hazard scenario. This is particularly obvious for multicollinear data (i.e., the correlation between a predictor variable and all other remaining variables in the model). Ridge regression remedies this problem by introducing a “penalty” term in the least-squares estimation of the regression coefficients. This penalty term can be thought of as a constraint on regression coefficients. Various methods have been developed to determine the appropriate value for the regularization parameter – in the present study, an extension of the generalized cross-validation (GCV) method called Robust GCV is used [5].

The functional form of the seismic demand model for the 6-story frame model selected for the present study includes the spectral acceleration values at 1st, elongated 1st, and 2nd mode periods, as follows:

\[
E[\ln(MIDR)] = \beta_0 + \beta_1 \ln[S_a(T_1)] + \beta_2 \ln[S_a(T_2)] + \beta_3 \ln[S_a(1.4T_1)]
\]  

(1)

\(MIDR\) refers to the maximum interstory drift demand in the structure and \(\beta_0 \sim \beta_3\) are the regression coefficients. Results of the simulations expressed in terms of peak demands versus the three selected intensity measures using the 2 methods are shown in Figure 2.

![Figure 2. Peak seismic demands versus selected intensity measures.](image)

Next, several realizations consisting of 10 randomly selected records were used to regenerate demand models for each of the two methods. The regression coefficients in each case for the two approaches are listed in Table 1. It is seen that the regression coefficients (with the exception of the constant term) vary significantly when using standard least squares regression whereas the coefficients estimated by Ridge regression remain stable for all three realizations.

Another measure of the performance of the demand model is to determine the mean error when the model developed with one data set is applied to another data set from the same hazard scenario. Model coefficients are first determined using a “training set” (consisting of 10 randomly selected records) and then applied to predict the response of a “validation set” (consisting of a different set of 10 records from the same original bin of 108 records). Results are displayed in Figure 3 which shows the mean error in the estimated versus observed demands using the two models. While both models have similar mean errors for the training set as expected, the errors increase almost two-fold for the standard least squares estimates, but is consistently below 15% for both datasets for Ridge regression.
Table 1. Regression coefficients for random realizations with 10 samples

<table>
<thead>
<tr>
<th></th>
<th>Least Squares</th>
<th>Ridge Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>Realization #1</td>
<td>-3.34</td>
<td>0.78</td>
</tr>
<tr>
<td>Realization #2</td>
<td>-3.51</td>
<td>0.15</td>
</tr>
<tr>
<td>Realization #3</td>
<td>-3.60</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Figure 3. Predictive capability of the two models: (a) Training set; (b) Validation set

**Conclusion**

Standard regression using least squares leads to overfitting the data with the result that the regression coefficients developed for one data set may not be applicable to another data set for the same hazard scenario. It is demonstrated that penalized least squares using Ridge regression overcomes this problem and is a more reliable method for developing demand models.

**References**