ICONHIC2019
2nd International Conference on Natural Hazards & Infrastructure
23-26 June, 2019, Chania, Greece

Selecting Building Characteristics to Predict Seismic Retrofit Costs of a Building Portfolio

J.F. Fung[1], S. Sattar, D.T. Butry, S. McCabe
National Institute of Standards and Technology

ABSTRACT

An accurate yet simple estimate of the retrofit cost plays an important role in the decision-making process of retrofitting existing buildings. Fung et al. (2018b) developed a predictive model to estimate seismic retrofit costs as a function of building characteristics such as building area, building age, and building model type. However, in practice, a decision maker may not have access to the full set of building characteristics required for estimating the retrofit cost, especially when dealing with a portfolio of buildings. Certain characteristics (e.g., building area) might be more readily available or easier to obtain than others (e.g., building type). This paper considers the tradeoff, in terms of prediction error, from not using all of the building characteristics necessary for prediction of retrofit cost. The results show that excluding certain characteristics from prediction, such as building type, lead to negligible increases in the prediction error. The paper also finds the minimal set of building characteristics needed to approximate the accuracy of the model that uses the full set of building characteristics. Findings of this study will help decisions makers to estimate retrofit costs without having to spend additional time and money to collect the full set of data on the building portfolio.

Keywords: seismic retrofit, retrofit cost prediction, GLM, regularization, variable selection

INTRODUCTION

One option for obtaining seismic retrofit cost estimates for a building is to hire an engineering consulting professional. However, even for a single building, the cost estimate requires detailed structural information, including the existing seismic detailing and material properties. Obtaining this information requires that the engineering professional examine the building on site. The process becomes time consuming and expensive as the number of buildings increases.

An alternative is to estimate retrofit costs for a building based on *historical* retrofit costs of buildings that have been retrofitted. Fung et al. (2018b) develop a predictive model to estimate seismic retrofit costs as a function of eight predictors, shown in Table 1. The advantage of this approach over hiring a professional is that the data required for prediction is generally easier to obtain and does not require on site inspections. Thus, retrofit cost predictions can be generated quickly and cheaply. On the other hand, such cost predictions may not be as accurate as those from an engineering professional.

In practice, some of the building characteristics presented in Table 1 may not be available for some or all buildings in a portfolio. Fung et al. (2018c), for instance, use the predictive modeling approach to obtain seismic retrofit cost estimates for a portfolio of buildings that does not include building age, height, or type.

---

[1] Corresponding Author: J.F. Fung, *National Institute of Standards and Technology, juan.fung@nist.gov*

Table 1. Target outcome, *Y*, and set of predictors, *X*, based on the retrofit cost model in Fung et al. (2018b).

Variable	Definition	Scale/Values
Y	Retrofit cost	Dollars per square foot
s	Seismicity	{Low, Medium, High, Very High}
p	Performance objective	{Life Safety, Damage Control, Immediate Occupancy}
b	Building group	{1, …, 8}[a]
Area	Total building area	Square feet
Age	Building age	Years
Height	Building height	Total above and below ground stories
Occup	Occupancy during retrofit	{Vacant; In-place; Temporarily relocated within building}
Historic	Building historic status	Is building deemed historic? (Yes or No)

[a] Building model types are categorized into eight building groups, as shown in Table 2.

Note that the performance objective, p, i.e., the target building performance after the retrofit, and the occupancy during retrofit, Occup, are the only predictors that correspond to retrofit *actions*. The other predictors are building characteristics. The performance objective categories represented in Table 1 are defined in FEMA (1994) as follows: Life Safety (LS) "allows for unrepairable damage as long as life is not jeopardized and egress routes are not blocked;" Damage Control (DC) "protects some feature or function of the building beyond life-safety, such as protecting building contents or preventing the release of toxic material;" and Immediate Occupancy (IO) "allows only minimal post-earthquake damage and disruption, with some nonstructural repairs and cleanup done while the building remains occupied and safe."

Two Motivating Questions

This paper explores the performance of the predictive model developed in Fung et al. (2018b) when some of the building characteristics are not available. The paper addresses two motivating questions:

1. What is the *minimal model*, that is, minimal in the number of predictors, for obtaining performance comparable to the *benchmark model* that includes all of the predictors?
2. What is the effect on performance from using a model that deliberately omits building age, height, and type (the *practical model*)?

To answer these questions, the paper compares three candidate models, the *benchmark model*, the *minimal model*, and the *practical model* by their performance in terms of *prediction error*. The next section describes the predictive modeling approach, and the following section presents the results.

METHODOLOGY

Fung et al. (2017, 2018a, 2018b) develop a predictive model for structural seismic retrofit costs, Y, as a function of the predictors shown in Table 1, $Y = f(X)$. The goal is to use historical data X to estimate a function $\hat{f}$ such that $\hat{Y} = \hat{f}(X_{new})$ is a reasonable prediction of retrofit costs for a new building with characteristics X_{new}.

This paper uses a Generalized Linear Model (GLM) to estimate *f*, as in Fung et al. (2018b).[2] Moran et al. (2007) discuss the use of GLMs for cost prediction. One key advantage of using GLMs, rather than standard linear regression models, for cost prediction is that results are easily interpretable in dollar terms (Fung et al., 2018b).

[2] The paper uses a GLM with gamma-distributed outcome and a log link. The "Linear" part of a GLM means f is a function of a linear combination of the predictors, $X\beta$, where β is a vector of coefficients; thus, $f(X) = e^{X\beta}$. For details, see Fung et al. (2018b).

Given an estimator $\hat{f}$, model performance is estimated using *prediction error*. This paper uses Root Mean Square Error (RMSE), given in Equation (1), as the measure of the prediction error,

$$RMSE(\hat{f}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{f}(X_i) - Y_i)^2}\ , \tag{1}$$

where Y_i is the actual retrofit cost for building *i* and $\hat{f}(X_i)$ is the predicted retrofit cost for building *i* and *m* is the number of observations. Note that both the actual cost and the predicted cost are in the scale of dollars per square foot and thus RMSE is easily interpretable on the same scale.

Prediction error is an important criterion for two distinct model development steps: *model selection* (choosing the best model out of a set of candidate models) and *model evaluation* (estimating a model's expected performance on new data, or "out-of-sample performance"). This paper uses *nested K-fold cross-validation* in order to perform both model selection and model evaluation (Fung et al., 2018b). The approach prevents the data used for selecting a model from contaminating the data used for model evaluation. This is especially useful when the number of observations used to fit, or *train*, the models is small.[3]

In order to address the two motivating questions, this paper considers three models:

1. The *benchmark model*, $\hat{f}_{benchmark}$, is the model that includes all of the predictors in Table 1.
2. The *minimal model*, $\hat{f}_{minimal}$, the model that includes the minimal number of predictors for obtaining performance close to that of the benchmark model.
3. The *practical model*, $\hat{f}_{practical}$, is the model that deliberately omits predictors that may not be easily available (in this case, building age, building height, and building type).

The three models are compared based on their out-of-sample prediction error. Thus, for instance, any increase in prediction error from omitting building age, height, and type can be interpreted as the cost of not collecting information on these predictors.

Regularization

In general, increasing the number of predictors in a regression model tends to improve performance on the *training data*, the data used to fit the model, though not necessarily on new data. This is known as *overtfitting* because the model tends to fit the training data precisely but cannot generalize when presented with new data.

The problem of choosing the best subset of predictors, e.g., those that minimize prediction error, is known as *variable selection*. It is worth noting that, since five of the eight predictors in Table 1 are *categorical*, the number of effective predictors used to estimate *f* is much larger than eight.[4] Rather than considering all possible combinations of the predictors (sometimes called stepwise variable selection), this paper uses *regularization* to find the minimal model, $\hat{f}_{minimal}$. Regularization penalizes model complexity (i.e., increasing the number of predictors) while minimizing a criterion such as prediction error (Zou and Hastie, 2005). Regularization is often used to prevent overfitting and can be used to enforce model simplicity.

In particular, the paper uses the lasso (least absolute shrinkage and selection operator) to estimate $\hat{f}_{minimal}$.[5] Lasso is especially useful for obtaining *sparse* solutions: the larger the penalty, the lower the number of *active* predictors used to train the model. Thus, regularization via lasso performs variable selection.

Not all regularization methods perform variable selection. For instance, ridge regression (Hoerl and Kennard, 1970) reduces the influence of correlated groups of predictors, but ultimately all predictors are used to train the model.[6] Lasso, on the other hand, does not account for correlations among predictors. The *elastic net* is a

[3] K-fold cross-validation splits the data into K mutually exclusive subsets then trains the model K times. At each iteration, one subset is used to estimate prediction error while the rest of the data is used to estimate $\hat{f}$. See Krstajic et al. (2014) for a review of the procedure as well as discussion of the potential pitfalls.

[4] In fact, Fung et al. (2018b) also include the combined effect of seismicity and the performance objective, an *interaction term*, that brings the total number of effective predictors to 35.

[5] Lasso penalizes the sum of the absolute values of the coefficients, $\sum_k|\beta_k|$, i.e., the L_1-norm of the coefficient vector, $||\beta||_1$ (Tibshirani, 1996).

[6] Ridge regression penalizes the sum of the squared values of the coefficients, $\sum_k {\beta_k}^2$, i.e., the squared L_2-norm of the

regularization method that combines lasso with ridge regression, thus performing variable selection while accounting for correlations among predictors (Zou and Hastie, 2005). Elastic-net regularization is a convex combination of lasso and ridge regression and therefore includes both as special cases.

In order for $\hat{f}_{minimal}$ (and $\hat{f}_{practical}$) to be compared against an appropriate benchmark, the benchmark model $\hat{f}_{benchmark}$ should be chosen as the optimal form of elastic-net regularization, which includes as a special case a model with no penalty (and hence no regularization). The model selection step of nested K-fold cross-validation is used to choose $\hat{f}_{benchmark}$.[7] The model evaluation step of nested K-fold cross-validation is then used to estimate the expected out-of-sample performance of each model.

The Training Data

The historical retrofit cost data used in this paper was originally collected for FEMA 156 (FEMA, 1994) and is freely available online. In particular, the data can be found as part of FEMA's archived Seismic Rehabilitation Cost Estimator (SRCE) software, (FEMA, 2013–2014).

The publicly available version of the data (the SRCE data) includes 1978 buildings, compared to the 2088 collected for FEMA 156. The SRCE data set is missing an important building characteristic that is used in FEMA 156: building occupancy class. Nevertheless, the discussion in FEMA 156 suggests this data set should be representative of commercial and residential buildings in the United States and Canada.

The SRCE data used for training includes all of the predictors shown in Table 1, as well as the structural seismic retrofit cost for each building. Building model types are categorized into eight *building groups*, as shown in Table 2.

Table 2. Building types, building groups, and their shares in the SRCE data.

Building Group	Building Type	Building Type Name	Share
1	URM	Unreinforced Masonry	30.08 %
2	W1 W2	Wood Light Frame Wood (commercial or industrial)	2.62 % 3.08 %
3	PC1 RM1	Precast Concrete Tilt Up Walls Reinforced Masonry with Metal or Wood Diaphragm	3.34 % 3.34 %
4	C1 C3	Concrete Moment Frame Concrete Frame with Infill Walls	6.75 % 16.64 %
5	S1	Steel Moment Frame	4.85 %
6	S2 S3	Steel Braced Frame Steel Light Frame	1.83 % 0.72 %
7	S5	Steel Frame with Infill Walls	7.01 %
8	C2 PC2 RM2 S4	Concrete Shear Wall Precast Concrete Frame with Infill Walls Reinforced Masonry with Precast Concrete Diaphragm Steel Frame with Concrete Walls	16.19 % 0.79 % 0.66 % 2.10 %

Table 3 presents a summary of statistical information from the SRCE data for the structural retrofit cost per square foot, as well as some of the predictors from Table 1. All cost and RMSE values in this paper are given in 2016 US dollars per square foot (1 ft = 0.3048 m). Further information on the data set can be found in Fung et al. (2018b).

coefficient vector, $||\beta||_2^2$ (Zou and Hastie, 2005).

[7] To be precise, α, the hyperparameter governing the convex combination of L_1 and L_2 penalties, is chosen using random grid search. For a given α, the penalty parameter, λ, is chosen based on search algorithms developed in Friedman et al. (2010). If $\lambda = 0$, then there is no penalty and thus no regularization. The best combination of α, λ is the best pair (in terms of RMSE) from the model selection step of nested K-fold cross-validation.

Table 3. Summary statistics for the outcome of interest, structural retrofit cost per square foot (1 ft = 0.3048 m) and select predictors in the training (SRCE) data, with N=1526 excluding Canadian buildings.

Variable	**Minimum**	**Mean**	**Median**	**Maximum**	**Standard Deviation**
Structural cost ($/sq ft)	0.49	36.03	23.33	675.42	44.74
Area (1000 sq ft)	0.15	68.98	28.67	1430.30	113.26
Age	2.00	44.29	40.00	153.00	22.13
Stories	1.00	3.12	2.00	38.00	2.99

RESULTS

This section presents the main results. First, the results from model selection and the resulting benchmark model. Second, the expected out-of-sample performance for the benchmark, minimal, and practical models.

Model Selection and the Benchmark Model

The benchmark model is the optimal model chosen in the model selection step: it is the model that minimizes prediction error in the model selection step of nested K-fold cross-validation. Table 4 presents the results of model selection and, hence, the resulting benchmark model. The value of $\alpha = 0.7$ implies elastic-net regularization that weighs lasso more heavily than ridge regression and, thus, favors a sparser model. The value of $\lambda = 0.002$, while not huge in magnitude, is sufficiently larger than zero that the regularization penalty is non-trivial.

Table 4. The benchmark model is defined as the model with optimal values of the regularization hyperparameters, α, λ based on nested K-fold cross-validation with K=10.

	$\boldsymbol{\alpha}$	$\boldsymbol{\lambda}$
Mean	0.707	0.002
Standard deviation	0.230	0.001

Table 5 presents prediction error results from the model selection step of nested K-fold cross-validation for each of the candidate models. It is worth noting that Fung et al. (2018b) do not use regularization. Since the optimal model from the model selection step, $\hat{f}_{benchmark}$, does use some form of regularization, it is worth comparing the benchmark model to the original model with no regularization, $\hat{f}_{original}$.

Table 5. Model selection for each of the candidate models based on prediction error, RMSE, as well as the standard deviation of RMSE, σ_{RMSE}. The benchmark model is defined as the optimal model: the model with the lowest RMSE. All values in dollars per square foot (1 ft = 0.3048 m).

Model	**RMSE**	$\boldsymbol{\sigma_{RMSE}}$
Benchmark model, $\hat{f}_{benchmark}$	38.57	1.06
Original model, $\hat{f}_{original}$	38.89	1.06
Minimal model, $\hat{f}_{minimal}$	38.80	1.17
Practical model, $\hat{f}_{practical}$	40.11	1.06

The results suggest that the practical model, $\hat{f}_{practical}$, is not optimal. This is not surprising: if a decision maker has all of the information necessary for prediction available, it should be used. Regularization will ensure that the model does not overfit. Nevertheless, the standard deviation of the RMSE estimates suggest that the practical model is only marginally sub-optimal.

The results suggest that some form of regularization is optimal and, in particular, regularization that favors sparsity. Note that the minimal and original models are only marginally sub-optimal (in terms of RMSE) to the benchmark model. In fact, the standard deviation of the RMSE estimates suggests that both $\hat{f}_{minimal}$ and $\hat{f}_{original}$ could be optimal. Thus, a decision maker could reasonably choose either $\hat{f}_{benchmark}$, $\hat{f}_{minimal}$, or $\hat{f}_{original}$.

On the other hand, the minimal model $\hat{f}_{minimal}$ uses 24 of 35 predictors and is thus only marginally sparser than the benchmark model $\hat{f}_{benchmark}$, which uses 25 of 35 predictors. In both cases, the models ignore several of the interactions between seismicity and the performance objective, suggesting that the other predictors do a good job of capturing this effect. Interestingly, both models discard the effect of the Damage Control performance objective, most likely because its effect on cost is correlated with the Life Safety and Immediate Occupancy performance objectives. The most noteworthy result of shrinkage is that the effect of building area as a predictor of cost is on the order of 10^{-6}, almost (though not identically) zero.

Model Performance

This section presents the estimates of expected out-of-sample performance for each of the candidate models (including the original model with no regularization). Expected out-of-sample performance is an estimate of prediction error on *new data*, obtained from the model evaluation step. The results are presented in Table 6.

Table 6. Estimated out-of-sample performance, RMSE, and standard deviation of RMSE, σ_{RMSE}, for each of the candidate models. All values in dollars per square foot (1 ft = 0.3048 m).

Model	RMSE	σ_{RMSE}
Benchmark model, $\hat{f}_{benchmark}$	38.61	8.97
Original model, $\hat{f}_{original}$	37.72	8.91
Minimal model, $\hat{f}_{minimal}$	37.69	8.91
Practical model, $\hat{f}_{practical}$	39.15	8.15

The main implication is that the minimal model, $\hat{f}_{minimal}$, dominates the benchmark model, $\hat{f}_{benchmark}$, in terms of performance on new data. Indeed, the minimal model outperforms all the candidate models, including the original model, $\hat{f}_{original}$. However, note that the standard deviation of RMSE estimates is much larger when estimating out-of-sample performance than when selecting a model (as shown in Table 5). Thus, the differences between RMSE estimates presented in Table 6 are not statistically significant, meaning the minimal model is no worse than the benchmark model.[8]

Most importantly, note that while the practical model, $\hat{f}_{practical}$, has the worst expected performance (as shown by the largest RMSE value in Table 6), the large variance in prediction error estimates suggests the practical model could achieve performance comparable to the benchmark or minimal models. In other words, the results suggest that there is no statistically significant penalty to deliberately omitting (or not collecting) information on building age, building height, and building type.

[8] For instance, application of Welch's t-test shows the difference in RMSE estimates for the benchmark and minimal models has a p-value of 0.823 and, thus, is far from statistically significant.

CONCLUSIONS

This paper considers an "optimal" model for predicting structural seismic retrofit costs and considers the impact on expected out-of-sample performance (prediction error on new data) when the model does not include all of the available predictors. The results suggest that, in fact, the optimal model itself does not require all of the predictors. Moreover, the paper considers a model that deliberately omits information on building age, building height, and building model type, as a thought experiment on what happens when a decision maker cannot (or does not want to) obtain this information for a portfolio of buildings. The results suggest there is a negligible penalty for omitting this information from the model. In fact, the results suggest all models achieve statistically indistinguishable performance on new data.

It would be interesting to compare performance of the GLM predictive model used in this paper with other models that can capture more nonlinearities (e.g., random forests, gradient boosting machines, and deep neural networks). This is left for future work.

Disclaimer

NIST policy is to use the International System of Units (metric units) in all its publications. In this paper, however, information is presented in U.S. Customary Units (inch-pound), as this is the preferred system of units in the U.S. earthquake engineering industry.

REFERENCES

FEMA. Typical Costs for Seismic Rehabilitation of Existing Buildings, Volume 1: Summary. *FEMA 156*, Federal Emergency Management Agency, 1994.

FEMA. SRCE: Seismic Rehabilitation Cost Estimator. https://www.fema.gov/media-library/assets/documents/30220, 2013–2014.

Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 2010, 33:1 1-22.

Fung JF, Butry DT, Sattar S, McCabe S. A Methodology for Estimating Seismic Retrofit Costs. *National Institute of Standards and Technology Technical Note*, 2017, NIST TN 1973.

Fung JF, Sattar S, Butry DT, McCabe S. Cost Estimates for the Seismic Retrofit of Federally Owned and Leased Buildings. *11th National Conference on Earthquake Engineering*, Los Angeles, 2018a.

Fung JF, Sattar S, Butry DT, McCabe S. A Predictive Modeling Approach to Estimating Seismic Retrofit Cots. *Earthquake Spectra* (under review), 2018b.

Fung JF, Sattar S, Butry DT, McCabe S. Estimating Structural Seismic Retrofit Costs for Federal Buildings. *National Institute of Standards and Technology Technical Note*, 2018c, NIST TN 1996.

Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 1970, 12:1 55-67.

Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics*, 2014, 6:10.

Moran JL, Solomon PJ, Peisach AR, Martin J. New Models for Old Questions: Generalized Linear Models for Cost Prediction. *Journal of Evaluation in Clinical Practice*, 13: 381-389.

Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 1996, 58: 267-288.

Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, 2005, 67: 301-320.