

Iowa State University

From the Selected Works of Cameron A. MacKenzie

2013

Empirical data and regression analysis for estimation of infrastructure resilience with application to electric power outages

Cameron A. MacKenzie, *University of Oklahoma Norman Campus*

Kash Barker, *University of Oklahoma Norman Campus*



Available at: https://works.bepress.com/cameron_mackenzie/3/

EMPIRICAL DATA AND REGRESSION ANALYSIS FOR ESTIMATION OF INFRASTRUCTURE RESILIENCE, WITH APPLICATION TO ELECTRIC POWER OUTAGES

Cameron A. MacKenzie¹, Kash Barker²

1 Doctoral candidate, School of Industrial and Systems Engineering, University of Oklahoma, 202 W. Boyd St., Room 124, Norman, OK 73019. Email: cmackenzie@ou.edu

2 Assistant Professor, School of Industrial and Systems Engineering, University of Oklahoma, 202 W. Boyd St., Room 124, Norman, OK 73019. Email: kashbarker@ou.edu

ABSTRACT

Recent natural disasters have highlighted the need for increased planning for disruptive events. Forecasting damage and time that a system will be inoperable is important for disruption planning. The resilience of critical infrastructure systems, or their ability to recover quickly from a disruption, can mitigate adverse consequences of the disruption. This paper quantifies the resilience of a critical infrastructure sector through the Dynamic Inoperability Input-Output Model (DIIM). The DIIM, which describes how inoperability propagates through a set of interdependent industry and infrastructure sectors following a disruptive event, includes a resilience parameter that has not yet been adequately assessed. This paper provides a data-driven approach to derive the resilience parameter through regression models. Data may contain different disruption scenarios, and regression models can incorporate these scenarios through the use of categorical or dummy variables. A mixed effects model offers an alternate approach of accounting for these scenarios, and these models estimate parameters based on the combination of all scenarios (fixed effects) and an individual scenario (random effects). These regression models are illustrated with electric power outage data and a regional disruption that uses the DIIM to model production losses in Oklahoma following an electric power outage.

KEYWORDS

Resilience, interdependence, regression, power outages

INTRODUCTION

Recent natural disasters, including the 2011 earthquake and tsunami in Japan, the 2010 earthquake in Haiti, and the 2005 hurricanes in the U.S. Gulf of Mexico, highlight the need for planning and preparation for disruptive events. While protection and prevention activities can mitigate the impact of disruptive events, governments and industries can never be certain of protecting against everything, especially events like the natural disasters. Preparedness efforts also require planning for response and recovery.

We define resilience as the ability of a system—be it a firm, industry, or critical infrastructure—to bounce back from a disruption. Differing views on the definition of resilience describe it as the post-event outcome of pre-event preparedness planning (Haimes et al. 2008) and as the “inherent and adaptive” ability to avoid potential losses (Rose 2004). The quantitative approach described in this paper follows the former definition, and we suggest that risk management planning can enhance recovery.

A number of modeling approaches have been proposed to measure resilience. Bruneau et al. (2003) quantify resilience by measuring the expected loss due to an earthquake over time. Zobel (2011) builds on this resilience metric by relating recovery time to the initial impact and analyzing how preparedness efforts might need to trade off reducing the initial impact with reducing recovery time. Other approaches to quantifying resilience include comparing the estimated loss in economic output with the maximum predicted loss in output (Rose 2007), developing an influence or decision diagram that incorporates pre-disruption and post-disruption decisions (McDaniels et al. 2008), calculating the ability of a node to remain connected to a degraded network (Dueñas-Osorio et al. 2004), and combining the ability of a network to provide service with the time to restore service when failure occurs (Whitson and Ramirez-Marquez 2009).

This work takes a different approach to calculating resilience by focusing on a data-driven quantification of the resilience coefficient from the Dynamic Inoperability Input-Output Model (DIIM) (Lian and Haimes 2006). The DIIM evaluates the initial impact of a disruptive event, the propagation of that adverse impact to several interdependent industries and infrastructure systems, and the ultimate recovery of the entire interconnected system. The DIIM requires a resilience coefficient for each industry or infrastructure system in the economy, and this resilience coefficient describes each industry’s ability to recover from a disruption. Despite a number of applications, the DIIM’s resilience coefficient has not been estimated either by modeling observational or empirical data.

Data-driven regression models can quantify the resilience coefficient for an individual industry or infrastructure system. Regression is a popular and successful method to quantify relationships among continuous variables, and the form of the DIIM lends itself to a regression model. An industry’s resilience may also depend on specific disruptive events, and regression models based on the DIIM can be expanded to account for different events. In this paper, a disruptive *event* like an electric power outage leads to inoperability in one or more industries or infrastructure systems, and several *scenarios* (e.g., winter storm, fire) can initiate the disruptive event. We focus our regression analyses on these different scenarios.

The paper begins by describing the foundational background to the DIIM and its resilience coefficient in Section 2. Section 3 details the regression models, which include a simple linear formula and two methods of accounting for different disruption scenarios: (i) categorical variables and (ii) a mixed effects model. Section 4 illustrates the regression approach with a data-driven case study dealing with electric power outages. Concluding remarks appear in Section 5.

DYNAMIC INOPERABILITY INPUT-OUTPUT MODEL AND ITS FOUNDATIONS

Several schema have been proposed for modeling interdependent industry and infrastructure sectors, including system dynamics models (Min et al. 2007), agent-based models (Bagheri and Ghorbani 2007), and network models (Dueñas-Osorio et al. 2007). Our risk-based interdependency modeling approach, whose methodological background is provided in this section, is an extension of the economic input-output model (Leontief 1936).

The Input-Output Model and Measuring Inoperability

The economic input-output model describes the flow of commodities among interconnected sectors of an economy [Leontief 1936]. The linear form of the input-output model is $\mathbf{x} = \mathbf{Ax} + \mathbf{c}$, which tracks the total production output of a set of n sectors or industries as a function of intermediate use by other sectors (\mathbf{Ax}) and final consumer demand (\mathbf{c}). Total output and final demand are both measured in dollars, and the $n \times n$ matrix \mathbf{A} quantifies the proportional requirements from other sectors to produce sector i 's output, for $i = 1, \dots, n$. Annually, the U.S. Census Bureau and the Bureau of Economic Analysis (BEA) undertake a significant data collection effort to collect commodity flow information, which is used to populate the \mathbf{A} matrix. Miller and Blair (2009) provide a good overview of input-output economics.

The Inoperability Input-Output Model (IIM) uses the Leontief input-output framework and commodity flow data to measure the interdependent effects from a disruption (Santos and Haimes 2004; Santos 2006). A decreased functional capability, or inoperability, in an infrastructure can propagate to a number of other physically and logically interconnected infrastructure or industry sectors. Eq. (1) presents the IIM. Vectors are of length n and matrices of size $n \times n$, where n is the number industry or sectors in the economy.

$$\mathbf{q} = \mathbf{A}^* \mathbf{q} + \mathbf{c}^* \Rightarrow \mathbf{q} = [\mathbf{I} - \mathbf{A}^*]^{-1} \mathbf{c}^* \quad (1)$$

The vector \mathbf{q} represents the inoperability vector. Each element measures the extent to which (in proportional form) a sector is not productive, and $q_i = (x_i - \hat{x}_i)/x_i$, where x_i is the production of sector i under normal circumstances and \hat{x}_i is the reduced level of production due to a disruption.

In a physical sense, \mathbf{q} represents the extent to which a sector is not functioning relative to its intended output. Inoperability is analogous to “unreliability” in the reliability engineering literature (Modarres et al. 2010), and can be linked with physical and economic interdependencies (Santos et al. 2007). The driver of inoperability is the demand perturbation vector expressed by \mathbf{c}^* , and $c_i^* = (c_i - \hat{c}_i)/x_i$, where c_i is the final demand for sector i under normal circumstances and \hat{c}_i is reduced demand. The matrix \mathbf{A}^* is the normalized interdependency matrix and describes the degree of interdependence among industry and infrastructure sectors. An element of \mathbf{A}^* indicates the proportion of additional inoperability that

the column sector contributes to the row sector. Like the \mathbf{A} matrix in the traditional input-output model, the \mathbf{A}^* matrix is derived from the BEA commodity flow data, and $\mathbf{A}^* = [\text{diag}(\mathbf{x})]^{-1}\mathbf{A}[\text{diag}(\mathbf{x})]$. Ultimately, the IIM provides an accounting for how a disruption that adversely impacts sector i can lead to inoperable conditions in other industry and infrastructure sectors.

Dynamic Inoperability Input-Output Model

The DIIM extends the IIM to model how inoperability propagates and dissipates over time (Lian and Haimes 2006). The original continuous-time DIIM, inspired by the Leontief (1970) dynamic input-output model is found in Eq. (2). Eq. (3) provides the discrete-time form.

$$\dot{\mathbf{q}}(t) = \mathbf{K}[\mathbf{A}^*\mathbf{q}(t) + \mathbf{c}^*(t) - \mathbf{q}(t)] \quad (2)$$

$$\mathbf{q}(t + 1) = \mathbf{q}(t) + \mathbf{K}[\mathbf{A}^*\mathbf{q}(t) + \mathbf{c}^*(t) - \mathbf{q}(t)] \quad (3)$$

Definitions of $\mathbf{q}(t)$, \mathbf{A}^* , and $\mathbf{c}^*(t)$ are the same as those for their counterparts in the IIM, except that $\mathbf{q}(t)$ and $\mathbf{c}^*(t)$ describe those values at a specific time t . \mathbf{K} is a diagonal matrix of resilience coefficients k_1, \dots, k_n and quantifies the ability of industry i to recover from a disruption. Greater k_i values correspond to quicker recoveries by an industry. MacKenzie et al. (2011) offers a slightly different interpretation, where \mathbf{K} describes how quickly the economy reaches equilibrium as determined by Eq. (1), and such an interpretation works especially well for a production environment.

The IIM and DIIM have been successfully applied to describe a number of risk-based applications, including multimodal transportation disruptions (Pant et al. 2011; MacKenzie et al. 2011), workforce losses (Barker and Santos 2010a; Orsi and Santos 2010), inventory management (Barker and Santos 2010b), and supply shortages (Xu et al. 2011). Barker and Haimes (2009) explore the impact of uncertainty in some of the DIIM's parameters.

Analyses using this model often assume that $\mathbf{c}^*(t) = 0$, which means that the disruption did not perturb final demand. In this case, inoperability results exclusively from the initial impacts. Under this assumption, an exponential model, as shown in Eq. (4), serves as a closed-form solution to the dynamic system of Eq. (2). In this exponential form, \mathbf{I} is the identity matrix and $\mathbf{q}(0)$ is the initial inoperability caused by the disruption at time $t = 0$.

$$\mathbf{q}(t) = \mathbf{q}(0)e^{-\mathbf{K}(\mathbf{I}-\mathbf{A}^*)t} \quad (4)$$

The scalar form of Eq. (4) is provided in Eq. (5) for a single industry i . The diagonal entry of \mathbf{A}^* , a_{ii}^* , measures the dependency of an industry on itself.

$$q_i(t) = q_i(0)e^{-k_i(1-a_{ii}^*)t} \quad (5)$$

If the initial inoperability $q_i(0)$ and the final desired inoperability $q_i(T_i)$ at recovery time T_i are known for a disruption, we can rearrange Eq. (5) and solve for the resilience coefficient k_i .

$$k_i = \frac{\ln[q_i(0)/q_i(T_i)]}{T_i} \left(\frac{1}{1 - a_{ii}^*} \right) \quad (6)$$

As Eq. (6) suggests, an accurate estimate of the resilience coefficient matrix \mathbf{K} for all n sectors requires knowledge about or data describing the initial inoperability, the final desired inoperability, and recovery time for each sector or industry. If these values are known, the DIIM can serve as a useful risk management tool to help policy makers quantify the economic consequences of potential disruptions. Assessing values for these parameters, especially T_i , poses a challenge, as the time to recover from a disruption is usually unknown. For example, state emergency managers may be interested in understanding the state-wide economic impacts of an extended power outage due to a natural disaster, but without a good estimate of either the time it takes to recover or the resilience coefficient, the DIIM may not provide accurate results.

Historical data describing disruptions for an individual industry i can help modelers overcome this challenge. As the next section demonstrates, Eq. (6) can be rearranged so that the model is linear in $\ln[q_i(0)]$ and T_i . Data on the initial inoperability $q_i(0)$ and the time until recovery T_i for specific disruptions can be incorporated into linear regression models in order to provide an estimation of the resilience coefficient k_i for a single industry. The next section describes the regression models and how the results can be used to estimate k_i .

QUANTIFYING THE RESILIENCE COEFFICIENT

The models in this section rely on m observations, where each observation consists of T_i and the initial inoperability $q_i(0)$. Regression analysis on these m observations allow us to estimate two parameters from Eq. (5): $k_i(1 - a_{ii}^*)$, where a_{ii}^* is known from the BEA data, and $q_i(T_i)$, where T_i is the known time it takes for sector i to recover to a desired level from a disruption. Full recovery implies that $q_i(T_i) = 0$, meaning that $T_i = \infty$ in accordance with the exponential model. Because this is impossible in reality, we expect our regression results to return estimates of $q_i(T_i)$ that are close to 0, or on the order of 0.01.

Simple Linear Regression Models

Different linear regression models can be used to estimate the two parameters. We first rearrange Eq. (6) in two different ways, shown in Eq. (7) and (8). In Eq. (7), $-1/k_i(1 - a_{ii}^*) \ln[q_i(T_i)]$ is the intercept term and $1/k_i(1 - a_{ii}^*)$ is the slope term. In Eq. (8), $\ln[q_i(T_i)]$ and $k_i(1 - a_{ii}^*)$ are the intercept and slope terms, respectively.

$$T_i = -\frac{1}{k_i(1 - a_{ii}^*)} \ln[q_i(T_i)] + \frac{1}{k_i(1 - a_{ii}^*)} \ln[q_i(0)] \quad (7)$$

$$\ln[q_i(0)] = \ln[q_i(T_i)] + k_i(1 - a_{ii}^*)T_i \quad (8)$$

Eq. (7) is perhaps more intuitive and reflects the natural cause and effect of a disruption, where a disruption causes an initial inoperability $q_i(0)$, and the time until recovery T_i depends on that initial inoperability.

Eq. (8), however, has some benefits that the Eq. (7) does not. First, the intercept and slope in Eq. (8) have no common terms, whereas Eq. (7) has a correlation between the intercept and slope

because both contain $1/k_i(1 - a_{ii}^*)$ terms. Assuming the slope and intercept are independent of each other will be important in the mixed effects model. Second, performing linear regression on Eq. (7) with our example data results in a statistically significant lack of fit for repeated observations of the independent variable, $\ln[q_i(0)]$. A statistically significant lack of fit for repeated observations indicates that the regression model should not be used (Draper and Smith 1998). The lack of fit is not statistically significant when the independent variable is T_i as in Eq. (8). Finally, k_i does not appear in a denominator in Eq. (7) as it does in Eq. (8), making it easy to calculate a point estimate, variance, and confidence interval for k_i after performing regression on Eq. (7).

Scenario-Specific Regression Models

A dataset may be divided into different categories, such as different disruption scenarios (e.g., natural events such as a tornado, earthquake, or hurricane), which may affect both the time and the initial inoperability. The parameters $q_i(0)$ and T_i may vary from scenario to scenario because different infrastructure degradations can result from different disruptive scenarios and result in different trajectories of recovery. The resilience of an industry or infrastructure system can be viewed as specific to a particular scenario, and k_i may most appropriately be calculated for a specific scenario. Dummy variables can be created to account for these categories. For a model with p categories, we create an $m \times (p - 1)$ matrix \mathbf{Z} . The columns of \mathbf{Z} in Eq. (9) correspond to the scenario categories, where the element (r,s) is 1 if the r th observation belongs to scenario category s and 0 otherwise. $\boldsymbol{\beta}$ is a regression vector of length $(p - 1)$.

$$\ln[q_i(0)] = \ln[q_i(T_i)] + k_i(1 - a_{ii}^*)T_i + \mathbf{Z}\boldsymbol{\beta} \quad (9)$$

Eq. (9) generates a different intercept term for each scenario category, but the slope remains the same for each category. Adding an interaction term between the time and dummy variables, as shown in Eq. (10), establishes a model where the slope changes for each category (Faraway 2006). Vector $\boldsymbol{\lambda}$ is of length $(p - 1)$ and \mathbf{T} is a $m \times m$ diagonal matrix and the diagonal entries are the entries in the vector T_i .

$$\ln[q_i(0)] = \ln[q_i(T_i)] + k_i(1 - a_{ii}^*)T_i + \mathbf{Z}\boldsymbol{\beta} + \mathbf{TZ}\boldsymbol{\lambda} \quad (10)$$

Mixed Effects Model

The final model we deploy is a mixed effects model. Such a model is useful when the data have clusters or categories and where variation in the model comes from two sources: within clusters and between clusters (Demidenko 2004). In the data used to estimate the resilience coefficient, different disruption scenarios can serve as clusters, and the mixed effects model can be viewed as an alternative to adding dummy variables. As a hierarchical or multi-level model, the mixed effects model assumes each cluster of data derives from an underlying distribution as opposed to separating the data like a linear model with dummy variables. The slope parameter $k_i(1 - a_{ii}^*)$ in the mixed effects model is the average or most likely estimate of all scenarios, but an individual scenario also has a random parameter that can be added to the average slope for that scenario.

Eq. (11) shows the mixed effects model where the superscript $j = 1, \dots, p$ refers to the scenario category. $\tilde{\mathbf{T}}^j$ is an $m^j \times 2$ matrix, where the first column is a column of ones and the second

column is T_i^j , and m^j represents the number of data points in scenario category j . Vector \mathbf{b}^j is a random vector of length 2.

$$\ln[q_i^j(0)] = \ln[q_i(T_i^j)] + k_i(1 - a_{ii}^*)T_i^j + \tilde{\mathbf{T}}^j\mathbf{b}^j \quad (11)$$

The model assumes that for each scenario category j , $\mathbf{b}^j \sim N(0, \sigma^2\mathbf{D})$. \mathbf{D} is a 2 x 2 matrix, whose diagonal entries describe the variance of each component of the \mathbf{b}^j vector. We assume that the components of \mathbf{b}^j are uncorrelated, so the off-diagonal entries of \mathbf{D} are zero. Such an assumption explains the importance behind selecting a linear model where the intercept and slope are not correlated. The parameter σ^2 is the variance of the errors (i.e., the mean-squared error term). The overall variance of the mixed effects models is usually less than the variance of an ordinary least-squares model (Demidenko 2004).

Several methods exist to solve for the parameters in the mixed effects model. One method is maximizing the likelihood to solve for $\ln[q_i(T_i^j)]$, $k_i(1 - a_{ii}^*)$, and σ^2 and using those parameters with the least squares approach to solve for \mathbf{b}^j (Demidenko 2004). Under this model, each scenario generates its own intercept and slope. The intercept for category j is $\ln[q_i(T_i^j)]$ plus the first component of \mathbf{b}^j , and the slope for category j is $k_i(1 - a_{ii}^*)$ plus the second component of \mathbf{b}^j .

Several types of tests can be used to analyze the statistical significance of a mixed effects model and its parameters (Khuri et al. 1998). We conduct a relatively simple test where the null hypothesis is $\mathbf{D} = \mathbf{0}$ (Demidenko 2004). If the null hypothesis is true, $b_j = 0$ and the random effects parameters do not contribute significantly to the model.

ILLUSTRATIVE EXAMPLE: ELECTRIC POWER OUTAGE AND RECOVERY

The regression models described here to isolate a measure of infrastructure resilience are illustrated with a case study of electric power outages in the United States. Because of their importance to the economy and the availability of data, power outages have been frequently examined in the risk analysis and reliability literature (Liu et al. 2005, 2008; Han et al. 2009; Dueñas-Osorio and Vemurua 2009; Guikema et al. 2010). The analysis presented here differs from these previous studies because of our interest in modeling the resilience and recovery from a power outage.

The set of data used describes power outages in the United States from January 2002 to June 2009 (U.S. Energy Information Administration 2003, 2004, 2005, 2006, 2007, 2008, 2009a, b). The publicly available data fields include: the time and date of a power outage, the power company that suffers the outage, the state or states affected by the power outage, the number of customers who lost power, the type or cause of disturbance, the loss in megawatts, and the date and time that power was restored.

We calculate $q_i(0)$ as the number of customers without power divided by the total number of retail customers in the affected state(s) (U.S. Energy Information Administration 2009c). Although a more accurate calculation of $q_i(0)$ may be to use the total number of customers serviced by the power company in the denominator, that information is not easily attainable. Our

approximation of $q_i(0)$ can be problematic for larger states. An event may cause large a power outage in a specific city or region, and it may take several days or even weeks to restore power. However, in a large state with tens of millions of retail customers, the denominator is so large that the estimate of $q_i(0)$ is still rather small.

Alternatively, the total number of residents in the county or counties could serve as the denominator for $q_i(0)$. However, the data does not always reveal the impacted counties for an outage. Some outages record specific counties, other outages list the city (e.g., Chicago), other outages record a portion of the state (e.g., northern Illinois), and some outages record only the state. In order to remain consistent across the data, we choose the number of retail customers in the impacted state as the denominator for $q_i(0)$. While this choice increases the size of the dataset, the ability to model localized outages is lost.

A single event can cause multiple outages in the database, such as outages in 2005 due to Hurricane Katrina. If the same event causes several outages, the numerator in $q_i(0)$ is the sum of all customers without power across the multiple companies impacted by the same event. The denominator is the total number of customers in the affected state(s). The time of the outage is assumed to be the longest that any of these power companies took to restore power. After this consolidation, the dataset consists of 368 observations.

The resilience parameter k_i is industry specific and can potentially be different for each industry in the economy. Data on the duration of electric power outages provide an estimate of k_i for the utilities industry, which is the North American industry classification that encompasses the electric power sector. Estimating the resilience parameters for other industries, like agriculture, manufacturing, or oil and gas, requires data on the severity and length of disruptions in those specific industries. The parameter a_{ii}^* is on the order of 10^{-4} for the utilities industry in the United States (BEA 2010a, b). Thus, the estimate of the resilience parameter k_i for the utilities industry is equal to the slope estimate $k_i(1 - a_{ii}^*)$ as calculated from the regression models using the data on electric power outages.

Previous IIM explorations of electric power outages include those caused by a hypothesized high-altitude electromagnetic pulse scenario (Haimes et al. 2005) and the 2003 blackout in the Northeast United States (Anderson et al. 2007).

Results of Simple Linear Regression Models

The simple linear regression model in Eq. (8) was applied to the electric power outage data, and Table 1 provides the point estimate and 95% confidence intervals of the two regression parameters. Fig. 1 depicts the 368 observations in the outage dataset, along with the linear model. Although the regression model is significant, the R^2 term is only 0.0757.

Table 1. Simple Linear Regression Results with 368 Data Points

Regression parameter	Point estimate	Confidence interval
$\ln[q_i(T_i)]$	-4.53	[-4.82, -4.24]
$k_i(1 - a_{ii}^*)$	0.00810	[0.00519, 0.0110]

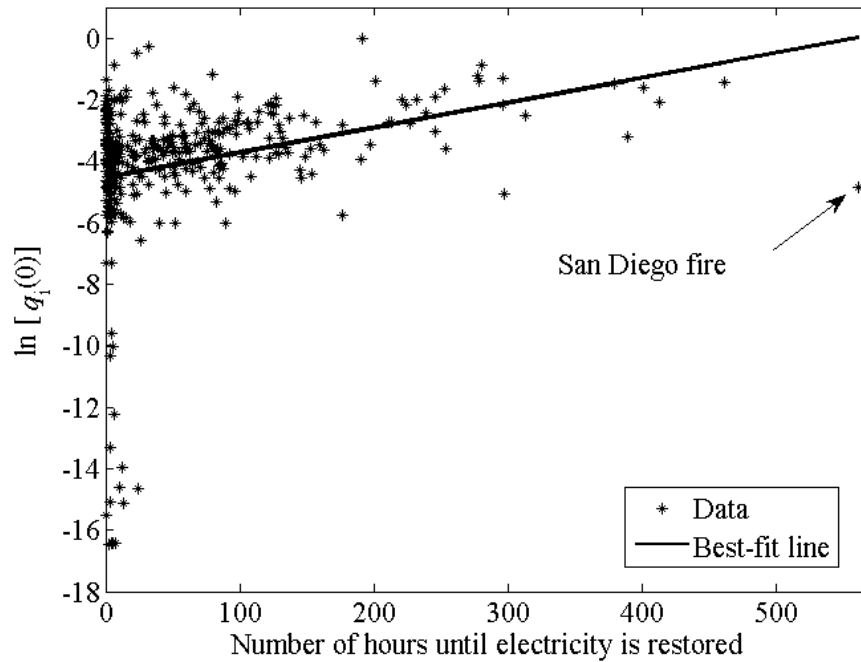


Fig. 1 Simple linear regression results with 368 recorded electrical outages

Checking the assumption of normally distributed error terms reveals that the residuals do not appear to follow a normal distribution (Fig. 2). In order to help correct the non-Gaussian distribution of the residuals and to improve the model's overall fit, we calculate Cook's statistic for each data point to check for outliers. Cook (1977) proposes a measure that calculates the influence of each data point based on the squared difference between the predicted values with all data points and the predicted values with all data points except for one (see Draper and Smith 1998).

According to Cook's statistic, the most influential data point corresponds to $T_i = 561$ and $q_i(0) = 0.008$. This data point represents a wildfire in San Diego in 2003 (see Fig. 1). The fire burned through over 200,000 acres. According to the database of outages (U.S. Energy Information Administration 2004) over 100,000 customers were without power, and power was not fully restored until three weeks later. Because California has over 13 million electric customers, $q_i(0)$ for this event is still very small. Cook's statistic for this data point is 0.30, and Cook's statistic for the second-most influential data point is 0.05. Because this data point has a large influence on the model relative to rest of the data, excluding this data point from the analysis on the grounds that it is an extreme outlier seems reasonable. Based on Cook's statistic, no other data point sufficiently influences the model to warrant exclusion.

Most of points that cause the overly broad tail on the left-hand side of the residual plot in Fig. 2 represent very small outages. The power companies quickly restored power after these outages. It is unlikely the DIIM would be deployed to analyze small outages like these because the outages would not impact the economy. From Fig. 1, power outages corresponding to $\ln[q_i(0)] > -8$ are separate from power outages corresponding to $\ln[q_i(0)] < -8$. The longest duration of a power outage corresponding to $\ln[q_i(0)] < -8$ is one day, which is a relatively

short time. Conversely, a few power outages corresponding to $-6 > \ln[q_i(0)] > -8$ lasted several days, and using a threshold for exclusion much greater than $\ln[q_i(0)] = -8$ appears unwise. Removing data points where $\ln[q_i(0)] < -8$ (equivalent to $q_i(0) < 0.00034$) eliminates relatively minor power outages and preserves all outages that lasted more than one day.

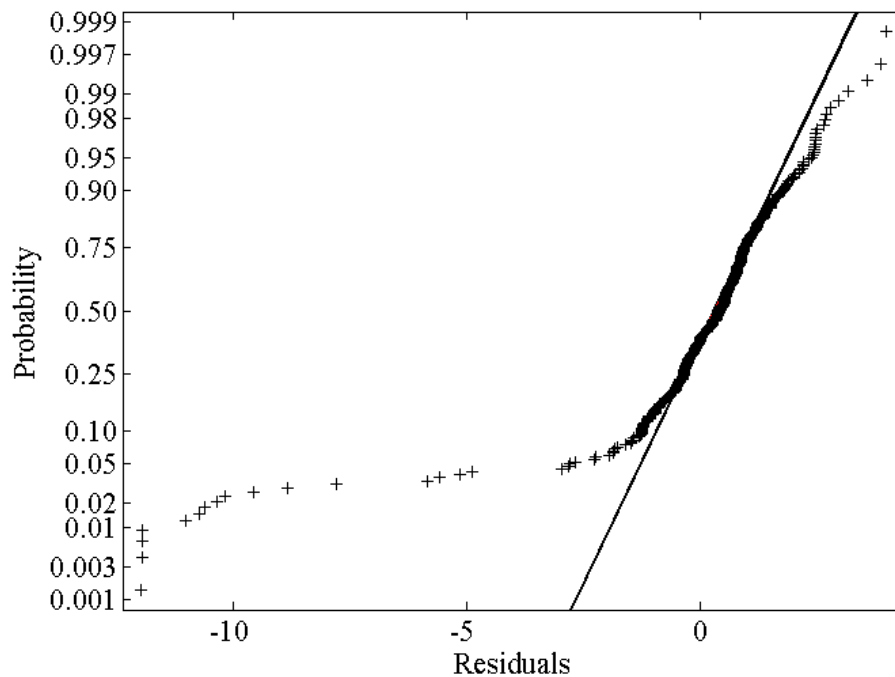


Fig. 2 Normal probability plot of residuals with 368 recorded electrical outages

With the removal of the San Diego fire data point and the data points describing small outages, the total number of observations is 352. With this revised dataset, we calculate the point estimates of the intercept and slope with 95% confidence intervals using ordinary least squares regression, as given in Table 2.

Table 2. Simple Linear Regression Results with 352 Data Points

Regression parameter	Point estimate	Confidence interval
$\ln[q_i(T_i)]$	-3.98	[-4.12, -3.83]
$k_i(1 - a_{ii}^*)$	0.00586	[0.00441, 0.00732]

With the removal of these data points, the intercept is closer to zero and the slope decreases. The estimate of the variance of errors is also greatly reduced (from 5.36 to 1.18), which results in tighter confidence intervals for the slope and intercept. For this revised model $R^2 = 0.153$. Although an improvement over the previous model, the small R^2 value demonstrates that this model still fails to explain most of the variation about the mean of the dependent variable, $\ln[q_i(0)]$.

We check the normality assumption of the residuals for this reduced dataset by visually examining a normality plot (Fig. 3) and conducting the Kolmogorov-Smirnov test (Law 2008).

Because the test does not reject the null hypothesis that the residuals follow a normal distribution, assuming that the residuals for this smaller dataset are normally distributed is appropriate.

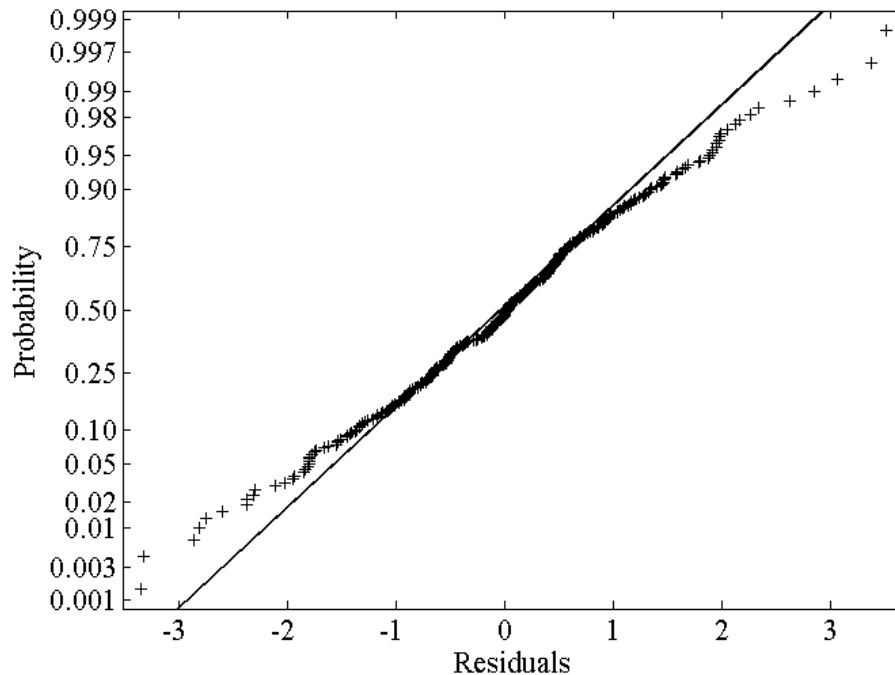


Fig. 3 Normal probability plot of residuals with 352 recorded electrical outages

Results of Scenario-Specific Regression Models

The cause of a power outage may impact the parameter estimates, and the Department of Energy lists approximately 100 different types or causes of outages in the dataset. We logically group these causes into six different scenarios: winter storm, hurricanes, equipment malfunction, voltage reduction (which usually means that demand exceeds supply), thunderstorm and tornado, and fire. Earthquake is another scenario, but the dataset only includes two earthquakes. Tornadoes are categorized with thunderstorms because the dataset only includes three tornadoes.

We assign five dummy variables as described in Eq. (9) to account for the six different scenarios. The intercept is initially the only term that changes for the different scenarios. Ordinary least squares regression is performed on the modified dataset where we remove the 16 data points as discussed earlier as well as the two earthquake-induced outages and another data point for which no cause was given. Point estimates with 95% confidence intervals are displayed in Table 3.

Using a partial F -test (Draper and Smith 1998), the dummy variable model is significant at the 0.05 level but not at the 0.01 level. The R^2 term improves slightly to 0.191.

Table 3. Linear Regression Results with Dummy Variables for Intercepts

Scenario	Regression parameter	Point estimate	Confidence interval
Winter storm	$\ln[q_i^1(T_i^1)]$	-3.68	[-4.02, -3.34]
Hurricanes	$\ln[q_i^2(T_i^2)]$	-3.59	[-4.07, -3.12]
Equipment malfunction	$\ln[q_i^3(T_i^3)]$	-4.55	[-4.94, -4.16]
Voltage reduction	$\ln[q_i^4(T_i^4)]$	-3.87	[-4.16, -3.59]
Thunderstorm and tornado	$\ln[q_i^5(T_i^5)]$	-3.97	[-4.15, -3.79]
Fire	$\ln[q_i^6(T_i^6)]$	-3.96	[-4.65, -3.27]
All outages	$k_i(1 - a_{ii}^*)$	0.00483	[0.00318, 0.00648]

When dummy variables are included, the slope parameter is less than the slope from the simple regression model with $n = 352$ data points, but the 95% confidence interval widens. The intercepts for the different scenarios have point estimates ranging from -3.5 to -4.6. The equipment malfunction scenario generates an intercept term that differs at the 0.05 significance level from the other scenarios except for the fire scenario. All other intercepts are not significantly different from each other. Based on this set of data, we cannot conclude definitively that the true value of $q_i(T_i)$ changes according to the scenario.

Because the slope parameter directly determines the resilience coefficient k_i , understanding if the scenario impacts the slope is more interesting than knowing its effect on the intercept. An interaction term between the dummy variables and time until recovery as described in Eq. (10) is added to the regression model, and Table 4 displays the results. We use the reduced dataset ($n = 349$).

Table 4. Linear Regression Results with Dummy Variables for Intercepts and Slopes

Scenario	$\ln[q_i^j(T_i^j)]$		$k_i^j(1 - a_{ii}^{*j})$	
	Point estimate	Conf interval	Point estimate	Conf interval
Winter storm	-3.45	[-3.93, -2.97]	0.00274	[-0.000761, 0.00625]
Hurricanes	-3.83	[-4.44, -3.22]	0.00642	[0.00331, 0.00952]
Equipment malfunction	-4.40	[-4.85, -3.96]	-0.0106	[-0.0328, 0.0117]
Voltage reduction	-3.77	[-4.08, -3.47]	-0.00723	[-0.0202, 0.00579]
Thunderstorm and tornado	-4.00	[-4.20, -3.80]	0.00545	[0.00307, 0.00782]
Fire	-3.45	[-4.67, -2.23]	-0.147	[-0.446, 0.151]

The results of this model raise questions of its validity and usefulness. The R^2 value only increases by 0.02 from the R^2 value of the model with a dummy variable for the intercept only. The partial F -test (Draper and Smith 1998) reveals that the model with a dummy variable for the intercept and slope model does not offer a statistically significant improvement over the model with a dummy variable for the intercept only.

Additionally, three scenarios (equipment malfunction, voltage reduction, and fire) generate negative slopes, which imply that inoperability worsens after the incident or that the power company has negative resilience. Negative slopes are unrealistic for the purpose of estimating k_i .

Because this model does not statistically improve over the previous model and it generates negative slope parameters for half of the scenarios, we conclude this model is not accurate enough to estimate the resilience parameter and exclude it from further analysis.

Results of Mixed Effects Model

Creating a hierarchical model like a mixed effects model can address the problem of negative slope estimates that is created by adding dummy variables for the slope parameter. The dummy variable model estimates the slope and intercept for a scenario based only on the data for that scenario, but the mixed effects model estimates the slope and intercept parameter for a scenario from the data for all scenarios (fixed effects) as well as the data for an individual scenario (random effects).

Building the mixed effects model as given in Eq. (11) with the reduced dataset ($n = 349$) reveals that the random effects parameters are insignificant (i.e., the null hypothesis that the random effect covariance matrix $\mathbf{D} = \mathbf{0}$ cannot be rejected). A mixed effects model on the full dataset except for the earthquake-induced data and the data point whose cause is unknown ($n = 365$) reveals that the variance of the random effect corresponding to the slope term is approximately zero, but the variance corresponding to the random effect of the intercept term is 1.19. Table 5 shows the point estimate and 95% confidence intervals of the two regression parameters from the mixed effects model.

Table 5. Mixed Effects Regression Results with 365 Data Points

Regression parameter	Point estimate	Confidence interval
$\ln[q_i(T_i)]$	-4.67	[-5.58, -3.77]
$k_i(1 - a_{ii}^*)$	0.00500	[0.00197, 0.00808]

As described in Demidenko (2004), a test statistic that follows the F distribution compares the sum of squares of the mixed effects model with that of the ordinary least squares regression. The p -value of the test statistic for this model is 8.15×10^{-11} , and the mixed effects model improves upon the simple linear regression model with 365 data points.

Because the variance for the slope parameter $k_i(1 - a_{ii}^*)$ is zero, it does not vary for different scenarios. Although the scenario or cause of a specific outage likely impacts a utility company's ability to restore electricity, this model's result suggests that knowing the scenario does not change our estimate of the resilience coefficient. The variance of the random effects parameter for $\ln[q_i(T_i)]$ indicates that the intercept for each scenario follows a normal distribution where the mean is the point estimate of the fixed effects parameter (-4.67) and the variance is 1.19.

Comparison of Results

This regression study was motivated by a desire to quantify the resilience parameter k_i that appears in the DIIM. Comparing the estimates of the slope parameter $k_i(1 - a_{ii}^*)$ generated by the different linear regression models can help us determine the best estimate for the resilience parameter. Out-of-sample validation also compares the different methods. We perform a 10-fold cross validation on each regression model in which we randomly divide the dataset into 10 equal subsets, calculate the best estimates for the two regression parameters based on 90% of the data, and test each model on the remaining 10%. We repeat this process 10 times and use a different

testing set for each run. The root-mean-square error (RMSE) and the mean absolute error (MAE) serve as performance metrics, as shown in Table 6.

Table 6. Comparison of Regression Results

Model	$k_i(1 - a_{ii}^*)$		Performance metrics		Expected amt of time (hours) to recover from $q_i(0) = 0.1$
	Point estimate	Confidence interval	RMSE	MAE	
Simple linear ($n = 368$)	0.00810	[0.00519, 0.0110]	2.17	1.29	275
Simple linear ($n = 352$)	0.00586	[0.00441, 0.00732]	2.19	1.22	286
Dummy variable for intercept only	0.00483	[0.00318, 0.00648]	2.16	1.21	347*
Mixed effects ($n = 365$)	0.00500	[0.00197, 0.00808]	2.10	1.30	451

* Because the intercept term is different for each dummy variable, the intercept parameter from the simple linear model ($n = 352$) is used to calculate time to recovery.

The performance metrics are similar for the four models. The simple linear ($n = 352$) and dummy variable models have the smallest MAE, but the RMSE for these two models is greater than the RMSE for the mixed effects model. Because these two models remove the extremely small $q_i(0)$ values when estimating the slope and intercept parameters, their estimates of $\ln[q_i(0)]$ are inaccurate when these small $q_i(0)$ values appear in a testing set. The RMSE highlights this inaccuracy. Because the MAE does not square the error, the linear model with the reduced dataset and the dummy variable model perform better according to MAE.

Table 6 also displays the expected number of hours it would take for electricity to be restored when the initial inoperability $q_i(0) = 0.1$ and given the point estimates of the parameters of each model. Because a_{ii}^* is very close to zero for the utilities industry, $k_i \approx k_i(1 - a_{ii}^*)$ for this industry.

The expected recovery time when $q_i(0) = 0.1$ ranges from 275 hours (11.5 days) to 451 hours (18.8 days). The expected recovery time can vary by a week depending on the model. This difference in recovery time is due to the slope as well as the intercept $\ln[q_i(T_i)]$. Because the mixed effects model includes additional data points for which the initial inoperability $q_i(0)$ is extremely small, this model's intercept is less than the intercept than the models with 352 data points, and the model estimates a longer recovery time. Because $k_i T_i = (\ln[q_i(0)] - \ln[q_i(T_i)]) / (1 - a_{ii}^*)$, the resilience coefficient and the recovery time are inversely proportional where $(\ln[q_i(0)] - \ln[q_i(T_i)]) / (1 - a_{ii}^*)$ serves as the constant of proportionality. Multiplying k_i by a positive real number α reduces the expected number of hours until electric power is restored by the factor $1/\alpha$. This demonstrates the sensitivity of the recovery time to the model's estimate of the resilience parameter.

Three models generate slopes between 0.00483 and 0.00586, and perhaps the best estimate is 0.00483 (from the linear model with the dummy variable for the intercept parameter). The

model's R^2 term of 0.191 is greater than those of the simple linear models, and the confidence interval is narrower than the mixed effects model. It also performs the best according to the MAE.

Case Study: Power Outage in Oklahoma

To illustrate the usefulness of this resilience parameter calculation, we model an electric outage due to a winter ice storm that occurred in the state of Oklahoma in December 2007. According to the Oklahoma Department of Emergency Management (ODEM) [2008], 634,749 customers or 33.5% of the state's retail customers lost power. Fig. 4 depicts the DIIM's estimate of inoperability in the utilities industry in Oklahoma and the fraction of Oklahoma customers without power that ODEM published at least daily and sometimes twice per day. The DIIM uses the slope estimate, $k_i(1 - a_{ii}^*) = 0.00483$, and confidence interval, $[0.00318, 0.00648]$, from the regression model with dummy variables as the resilience parameter.

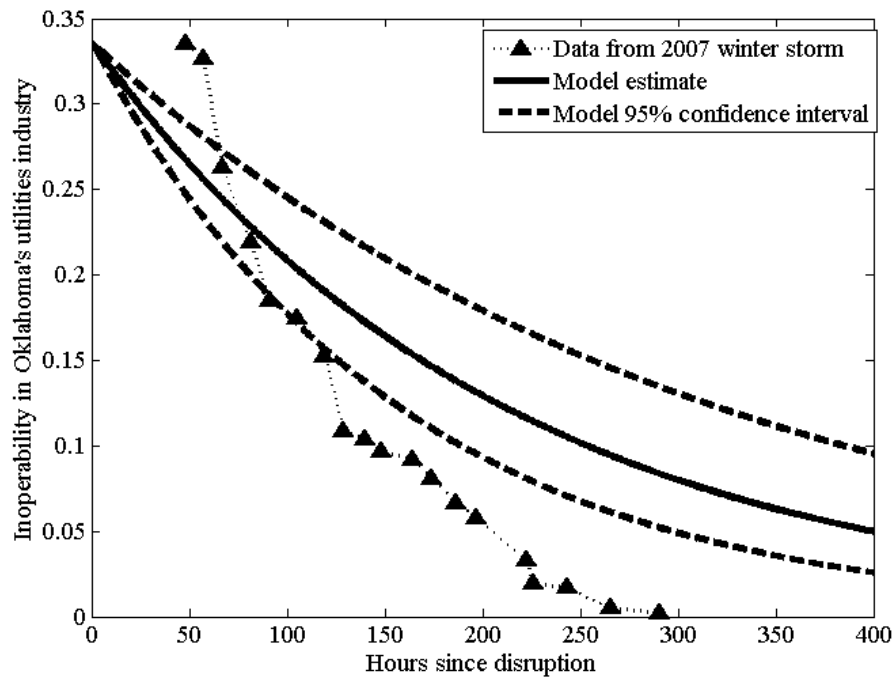


Fig. 4 Comparison of time to recovery between model estimate and historical data from 2007 winter storm in Oklahoma

Although the winter storm occurred on December 9, the peak number of customers without power was not reported until 48 hours later on December 11. The electric power companies seemed to arrive at an accurate count of the number of customers without power two days after the disruption occurred. In Fig. 4, the utilities industry's inoperability based on the historical data is 0.335 at time $T = 48$ hours. Because many of these customers likely suffered power loss when the winter storm struck Oklahoma on December 9, the model starts with an initial inoperability of 0.335 at $T = 0$.

The model overestimates the actual recovery time. The utilities industry achieved an inoperability equal to 0.017 in 244 hours (or 10.2 days), but the model predicts an inoperability

of 0.104 at that same time. Several reasons could explain the discrepancy between the model and the data. First, according to an Oklahoma utility company, the 2007 ice storm caused more power outages than any other storm in the state's history (Money and Bisbee 2007). Companies brought in employees from around the United States to restore power. This attention, combined with the fact that the storm was fairly localized to Oklahoma and did not impact surrounding states, may have enabled the utility companies to restore power more quickly than the model predicts. Second, 77% of the outages were concentrated in Oklahoma's two metropolitan areas, Oklahoma City and Tulsa. Restoring power in these two cities may have been easier than if the power outages had predominately occurred in rural areas.

The difference between the model's prediction and the data demonstrates the challenge of using a single number to model the recovery trajectory of a specific disruption. The trajectory of the actual recovery shows perhaps three different rates. Very little or no recovery may have occurred from $T = 0$ to $T = 48$ hours, but inoperability decreased rapidly from 0.335 at $T = 48$ hours to 0.108 at $T = 129$ hours. The rate of recovery slowed from $T = 129$ hours to $T = 223$ hours, and inoperability decreased to 0.033. The third recovery rate occurred while the remaining outages were repaired from $T = 223$ to $T = 291$ hours. A model that incorporates one resilience parameter for an industry cannot accurately capture the different rates of recovery.

These three different recovery rates echo the process of restoring electric power after an interruption, which often follows three phases (Clay Electric Cooperative 2007; Puget Sound Energy 2010). First, a utility company assesses the damage, which is usually accomplished within the first couple of days following the outage. In the 2007 winter storm, the first phase lasted 48 hours, which is the time when the peak number of customers without power was recorded. During the second phase, recovery occurs at a predictable rate until roughly 80% of power is restored, and phase three is more unpredictable as the utility company often repairs and replaces equipment for specific and localized populations. A piecewise treatment of the regression approaches to isolate the second and third phases is considered an interesting exploration of future work.

Given that our primary motivation behind modeling industry resilience is to quantify the economic impacts of disruptions, we use the estimate of resilience within the DIIM to calculate inoperability in all industries across the state of Oklahoma due to a disruption like the 2007 winter storm. BEA data (2010a, b) enables us to estimate \mathbf{A}^* for the state of Oklahoma as described in Section 2. The model includes 20 industries or economic sectors.

We assume the initial impact of the power outage is equivalent across all industries, and the initial inoperability is the fraction of customers without power: $q_i(0) = 0.335$ for all i . The slope parameter from the regression model, $k_i(1 - a_{ii}^*) = 0.00483$, determines the resilience parameter for the utilities industry, $i = 3$. Because $a_{3,3}^* = 8.04 * 10^{-4}$ for the utilities industry, $k_3 = 0.00483$. The previous modeling exercise provides an estimate for a single industry, utilities, but the DIIM requires resilience parameters for all 20 sectors. The other 19 industries suffer a power outage, but we assume that their facilities are not physically impacted. We assign a value of $k_i = 1$ for all $i \neq 3$ to represent a quick recovery for the non-utility industries.

Fig. 5 shows the recovery for these 20 industries as estimated by the DIIM with the above values for resilience and initial inoperability. The total output lost due to this power outage is \$104 million, with \$27.5 million of these losses occurring during the first 24 hours of the power outage. Each of the 19 non-utilities industries achieves an inoperability of less than 0.02 within 10 hours of the disruption. Two industries—mining, oil, and gas and transportation and warehousing—recover more slowly than the other industries because inoperability in utilities has the greatest interdependent impact on these industries. As determined by \mathbf{A}^* , inoperability in mining, oil, and gas equals 5.2% of inoperability in utilities, and inoperability in transportation and warehousing equals 3.4%. The inoperability in each of the remaining industries equals less than 1% of inoperability in the utilities industry.

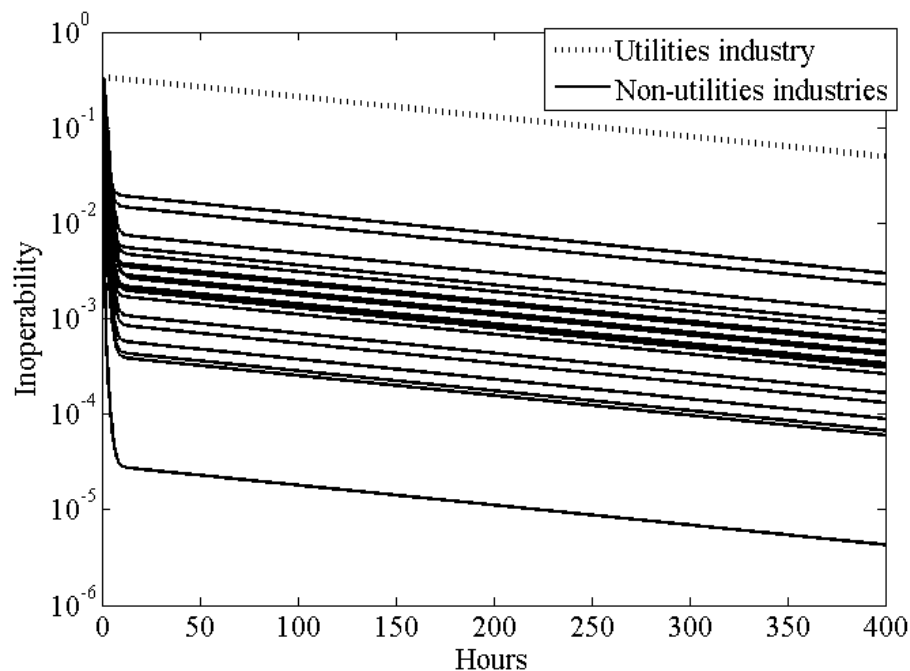


Fig. 5 Inoperability due to power outage in Oklahoma City with $k_3 = 0.00483$

If $k_i(1 - a_{ii}^*) = 0.00318$, the lost output in Oklahoma totals \$148 million. If the upper bound on the slope parameter is used, i.e., $k_i(1 - a_{ii}^*) = 0.00648$, lost output drops to \$82 million. The point estimate and bounds of the resilience coefficient form three different recovery scenarios although even the best-case scenario likely overestimates the output losses compared to the actual recovery data from the 2007 winter storm.

Table 7 depicts the lost output for each industry in Oklahoma for these three model-based recovery scenarios. The resilience parameter directly impacts the ability of the utilities industry to recover, and thus, the output loss in this industry is most sensitive to this parameter. Because the mining, oil, and gas industry is heavily impacted by the utilities industry, the lost output in this industry exceeds all other industries except for utilities. Manufacturing experiences a large output loss because this industry produces the most, and an initial inoperability of 0.335 results in millions of dollars of lost output for this industry.

Table 7. Lost Output per Industry

<i>i</i>	Industry	Lost output (millions of dollars)		
		$k_3 = 0.00318$	0.00483	0.00648
	All industries	148.40	104.04	82.28
1	Agriculture, forestry, and fishing	1.06	0.88	0.80
2	Mining, oil, and gas	28.03	19.22	14.90
3	Utilities	77.27	50.95	38.03
4	Construction	1.97	1.47	1.23
5	Manufacturing	15.68	12.18	10.46
6	Wholesale trade	1.96	1.57	1.37
7	Retail trade	0.89	0.79	0.74
8	Transportation and warehousing	6.24	4.38	3.46
9	Information	1.18	0.96	0.86
10	Finance and insurance	1.85	1.50	1.33
11	Real estate	2.64	2.15	1.90
12	Professional, scientific, and technical services	2.67	2.03	1.71
13	Company management	0.74	0.56	0.47
14	Administration and waste management	1.60	1.27	1.11
15	Educational services	0.07	0.06	0.06
16	Health care	0.63	0.62	0.62
17	Arts, entertainment, and recreation	0.12	0.10	0.10
18	Accommodation and food services	0.73	0.59	0.52
19	Other services, except public administration	0.53	0.46	0.42
20	Federal, state, and local government	2.55	2.31	2.19

Table 8 summarizes the key insights of applying the regression models to quantify resilience in the electric power sector. Removing outliers, removing very small power outages, and introducing variables to model specific scenarios improve the model's accuracy. Despite these additions and subtractions, the model fails to capture much of the variation in the data, a point to which we will return in the conclusion. With our estimate of the resilience coefficient, the DIIM predicts that a power outage in the state of Oklahoma would cost the state about a hundred million dollars in lost output. Utilities; mining, oil, and gas; manufacturing; and transportation and warehousing would experience the greatest impact.

CONCLUSION

The DIIM quantifies the economic losses due to the onset and propagation of a disruption, but one obstacle to its use has been estimating the resilience parameter that dictates the trajectory of sector recovery. This paper deployed a variety of data-driven regression models to address that obstacle and applied the models to different electric power disruption scenarios. We began with a simple regression model and attempted to achieve a better fit by eliminating irrelevant points, adding scenario-specific dummy variables and interaction terms, and testing for random effects through a mixed effects model.

Table 8. Key Insights of Applying Regression Models to Electric Power Outage Data

Modeling approach	Key insights
Simple linear regression	<ul style="list-style-type: none"> Model accuracy increases after removal of outliers and extremely small disruptions
Scenario-specific regression using dummy variables	<ul style="list-style-type: none"> Including scenarios for the intercept decreases error and increases adjusted R^2 value Model that includes scenarios for slope parameter generates unrealistic results (negative slopes)
Scenario-specific regression using mixed effects model	<ul style="list-style-type: none"> Slope parameter $k_i(1 - a_{ii}^*)$ has no random effects, which implies that the value of the resilience coefficient k_i is not impacted by the type of scenario
Case study of winter storm in Oklahoma	<ul style="list-style-type: none"> Model predicts a conservative time to recovery of utilities industry Output losses from a severe winter storm in Oklahoma would total about \$100 million

The dummy variables and the mixed effects model each incorporate individual scenario disruptions and measure the extent to which the industry resilience parameter changes for different scenarios. Both the intercept and slope describe the relationship between the initial impact and recovery time, often referred to as *robustness* and *rapidity* in the resilience literature (Bruneau et al. 2003; Zobel 2011). The intercept changes when the model includes different power outage scenarios, and we conclude that the power outage scenario impacts the relationship between the initial impact and recovery time. Because the slope term represents the industry resilience coefficient and remains constant for different scenarios, our estimate of k_i for the utilities industry is not scenario specific. Developing models that account for other aspects of recovery, such as the length of time to evaluate the cause of an outage, geographical and regional differences, and urban-versus-rural power outages could help increase the accuracy of the overall model and may produce resilience coefficients that are scenario specific.

The models generally produce statistically significant results although the R^2 terms are very small. The variation in the dataset leads to some difficulty in developing a general regression model describing recovery. Similarly, issues could arise from the DIIM's assumption of exponential recovery, and different DIIM extensions could relax this assumption. Further, our estimate of inoperability as a function of the total number of retail customers in a state could be flawed. We would prefer to estimate inoperability as a fraction of the total number of customers serviced by a power company if that data were available. Extracting the observations that give a specific location for the power outage, like a city or region, would enable inoperability to be based on the number of customers in that location. Such an extraction would result in a smaller database, and the results may not be more accurate than those we have already presented.

Perhaps any deterministic model would have difficulty in capturing the large variation that exists in the data. Probabilistic models could be deployed to model the recovery time. One approach would be to ignore the initial impact and fit a probability distribution to the observed recovery times by matching moments. Such a procedure should result in a probability distribution for the resilience coefficient although resilience would no longer be a function of the initial impact. Another approach is a hierarchical Bayes model (Gelman et al. 2004), which would assume that

recovery time follows a distribution (e.g., normal, exponential). One or more parameters of that distribution would follow another distribution that could be determined in part by the initial inoperability and scenario.

The case study of the winter storm in Oklahoma suggests another approach. Rather than using the initial impact and recovery time from a few hundred separate disruptions to estimate resilience, the resilience coefficient can be calculated from a single disruption where inoperability is recorded at multiple points in time during the same disruption. Performing regression on the data from the 2007 winter storm in Oklahoma produces an estimate of $k_i = 0.015$, which is an order of magnitude greater than the estimates based on final recovery times for hundreds of separate disruptions. Because this resilience estimate is based on a single disruption, using this value in the DIIM requires testing its applicability to other power outages. This resilience estimate does not account for the 48 hours that elapsed between the storm's arrival and the time when peak inoperability was recorded.

To our knowledge, this paper represents the first time the resilience coefficient in the DIIM has been empirically assessed. Future work can apply these regression models to data describing past disruptions in other sectors and infrastructure systems to quantify their resilience. Potential sectors include transportation, manufacturing, and service sectors like banking and finance. Such modeling efforts can provide greater clarity and information on the resilience coefficient matrix \mathbf{K} in the DIIM, enabling the DIIM to produce more accurate and meaningful results.

Even if the DIIM is not used as an economic model, the slope parameter in the regression models can represent the rate of recovery of an infrastructure system for an exponential model (Reed et al. 2009). Models that simulate the failure and restoration of complex systems can use these estimates to generate realistic scenarios (Luna et al. 2011). Understanding and predicting the rate of recovery of disabled or inoperable infrastructure and economic sectors can help decision makers examine the tradeoffs between the cost of protection and the consequences of different disruptions (Tsang et al. 2002). For example, preparedness strategies addressing system hardening could alter the value of $q_i(0)$, prepositioned recovery supplies could reduce T_i , and inventory and storage activities could affect $c_i^*(t)$. Because resilience is a function of these variables in the DIIM, we can quantify the impact of these preparedness strategies or decisions on the value of k_i . Modeling efforts like the one presented in this paper can lead to better risk management strategies and help policy makers prioritize among potential disruptions.

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation, Division of Civil, Mechanical, and Manufacturing Innovation, under award 0927299. We also thank the thoughtful suggestions of the reviewers, whose comments contributed significantly to this paper.

REFERENCES

Anderson, C. W., Santos, J. R., and Haines, Y. Y. (2007). "A risk-based input-output methodology for measuring the effects of the August 2003 northeast blackout." *Econ. Syst. Res.*, 19(2), 183-204.

- Bagheri, E. and Ghorbani, A. A. (2007). "Conceptualizing critical infrastructures as service oriented complex interdependent systems." *Proc., Int. Conf. on Inform. Tech. and Manage.*, Hong Kong.
- Barker, K. and Haimen, Y. Y. (2009). "Uncertainty analysis of interdependencies in dynamic infrastructure recovery: Applications in risk-based decision making." *J. Infrastruct. Syst.* 15(4), 394-405.
- Barker, K. and Santos, J. R. (2010a). "A risk-based approach for identifying key economic and infrastructure sectors." *Risk Anal.*, 30(6), 962-974.
- Barker, K. and Santos, J. R. (2010b). "Measuring the efficacy of inventory with a dynamic input-output model." *Int. J. Prod. Econ.*, 126(1), 130-143.
- Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W.A., and von Winterfeldt, D. (2003). "A framework to quantitatively assess and enhance the seismic resilience of communities." *Earthquake Spectra* 19(4), 733-752.
- Bureau of Economic Analysis (BEA). (2010a). "Make table, before redefinitions, 2008." *1998-2009 summary make annual I-O table before redefinitions*, U.S. Department of Commerce, (http://www.bea.gov/industry/io_annual.htm) (Sept. 2011).
- BEA. (2010b). "Use table, before redefinitions, 2008." *1998-2009 summary use annual I-O table before redefinitions*, U.S. Department of Commerce, (http://www.bea.gov/industry/io_annual.htm) (Sept. 2011).
- Clay Electric Cooperative, Inc. (2007). "Power restoration process." (<https://clayelectric.com/outage.aspx>) (Feb. 2012).
- Cook, R.D. (1977). "Detection of influential observation in linear regression." *Technometrics* 19(1), 15-18.
- Demidenko, E. (2004). *Mixed models: Theory and applications*, John Wiley & Sons, Hoboken, NJ.
- Draper, N. R. and Smith, H. (1998). *Applied regression analysis*, 3rd Ed., John Wiley & Sons, New York.
- Dueñas-Osorio, L., Craig, J. I., and Goodno, B. J. (2004). "Probabilistic response of interdependent infrastructure networks." *Proc., Annu. Meeting of the Asian-Pacific Network of Centers for Earthquake Eng. Research (ANCER)*, Honolulu, HI.
- Dueñas-Osorio, L., Craig, J. I., Goodno, B. J., and Bostrom, A. (2007). "Interdependent response of networked systems." *J. Infrastruct. Syst.*, 13(3), 185-194.

Dueñas-Osorio, L. and Vemurua, S. M. (2009). "Cascading failures in complex infrastructure systems." *Structural Safety*, 31(2), 157-167.

Faraway, J. J. (2006). *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*, Chapman & Hall, Boca Raton, FL.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*, 2nd Ed., Chapman & Hall, Boca Raton, FL.

Guikema, S. D., Quiring, S. M., and Han, S.-R. (2010). "Prestorm estimation of hurricane damage to electric power distributions." *Risk Anal.*, 30(12), 1744-1752.

Haimes, Y. Y., Horowitz, B. M., Lambert, J. H., Santos, J. R., Lian, C., and Crowther, K. G. (2005). "Inoperability Input-Output Model for interdependent infrastructure sectors. II: Case studies." *J. Infrastruct. Syst.*, 11(2), 80-92.

Haimes, Y. Y., Crowther, K. G., and Horowitz, B. M. (2008). "Homeland security preparedness: Balancing protection with resilience in emergent systems." *Syst. Eng.*, 11(4): 287-308.

Han, S.-R., Guikema, S. D., and Quiring, S. M. (2009). "Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models." *Risk Anal.* 29(10), 1443-1453.

Khuri, A. I., Mathew, T., and Sinha, B. L. (1998). *Statistical tests for mixed linear models*, John Wiley & Sons, New York.

Law, A. M. (2008). *Simulation modeling and analysis*, 4th Ed., Tata McGraw-Hill Publishing Company Limited, New Dehli.

Leontief, W. W. (1936). "Quantitative input and output relations in the economic system of the United States." *Rev. Econ. Stat.*, 18(3), 105-125.

Leontief, W. W. (1970). "The dynamic inverse." *Contributions to input-output Analysis*, A. P. Carter and A. Brody, eds., North-Holland Publishing, Amsterdam, 17-46. Reprint in W. W. Leontief (1986), *Input-output economics*, 2nd Ed., Oxford University Press, New York, 294-320.

Lian, C. and Haimes, Y. Y. (2006). "Managing the risk of terrorism to interdependent infrastructure systems through the Dynamic Inoperability Input-Output Model." *Syst. Eng.*, 9(3), 241-258.

Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R. (2005). "Negative binomial regression of electric power outages in hurricanes." *J. Infrastruct. Syst.*, 11(4): 258-267.

Liu, H., Davidson, R. A., and Apanasovich, T. V. (2008). "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms." *Rel. Eng. Syst. Safety*, 93(6), 897-912.

Luna, R., Balakrishnan, N., and Dagli, C. H. (2011). "Postearthquake recovery of a water distribution system: Discrete event simulation using colored petri nets." *J. Infrastruct. Syst.* 17(1), 25-34.

MacKenzie, C. A., Barker, K., and Grant, F. H. (2011). "Evaluating the consequences of an inland waterway port closure with a dynamic multiregional interdependency model." *IEEE Trans. Syst., Man, Cybern.—Part A: Syst. Humans*, in press. doi: 10.1109/TSMCA.2011.2164065

McDaniels, T., Chang, S., Cole, D., Mikawoz, J., and Longstaff, H. (2008). "Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaption." *Global Environmental Change*, 18, 310-318.

Miller, R. E. and Blair, P. D. (2009). *Input-output analysis: Foundations and extensions*, 2nd Ed., Cambridge University Press, Cambridge.

Min, H. J., Beyeler, W., Brown, T., Son, Y. J., and Jones, A. T. (2007). "Toward modeling and simulation of critical national infrastructure interdependencies." *IEE Trans.*, 39(1), 57-71.

Modarres, M., Kaminskiy, M., and Krivtsov, V. (2010). *Reliability engineering and risk analysis: A practical guide*, 2nd ed., CRC Press, Boca Raton, FL.

Money, J. and Bisbee, J. (2007). "Electricity: Outages could last for days." *The Oklahoman* (Oklahoma City), December 11, 8A.

Oklahoma Department of Emergency Management. (2008). "Winter weather event, December 9 – 10, 2007." (http://www.ok.gov/OEM/Emergencies_&_Disasters/2007/Winter_Weather_Event_20071209_-_Master/) (Dec. 2011).

Orsi, M.J. and Santos, J. R. (2010). "Estimating workforce-related economic impact of a pandemic on the Commonwealth of Virginia." *IEEE Trans. Syst., Man, Cybern.—Part A: Syst. Humans*, 40(2), 301-305.

Pant, R., Barker, K., Grant, F. H., and Landers, T. L. (2011). "Interdependent impacts of inoperability at multi-modal transportation container terminals." *Transportation Res. Part E: Logistics Transportation*, 47(5), 722-737.

Puget Energy Sound. (2010). "How power gets restored." Fact Sheet. (http://pse.com/aboutpse/PseNewsroom/MediaKit/3813_how_power_gets_restored.pdf) (Feb. 2012).

Reed, D. A., Kapur, K. C., and Christie, R. D. (2009). "Methodology for assessing the resilience of networked infrastructure." *IEEE Syst. J.*, 3(2), 174-180.

Rose, A. (2004). "Defining and measuring economic resilience to disasters." *Disaster Prevention Manage.*, 13(4), 307-314.

Rose, A. (2007). "Economic resilience to natural and man-made disasters: Multidisciplinary origins and contextual dimensions." *Environmental Hazards* 7, 383-398.

Santos, J. R. and Haimes, Y. Y. (2004). "Modeling the demand reduction input-output (I-O) inoperability due to terrorism of interconnected infrastructures." *Risk Anal.*, 24(6), 1437-1451.

Santos, J. R. (2006). "Inoperability input-output modeling of disruptions to interdependent economic systems." *Syst. Eng.* 9(1), 20-34.

Santos, J.R., Haimes, Y. Y., and Lian, C. (2007). "A framework for linking cyber security metrics to the modeling of macroeconomic interdependencies." *Risk Anal.*, 27(4), 1283-1297.

Tsang, J. L., Lambert, J. H., and Patev, R. C. (2002). "Extreme event scenarios for planning of infrastructure projects." *J. Infrastruct. Syst.* 8(2), 42-48.

U.S. Energy Information Administration. (2003). "B2: Major disturbances and unusual occurrences, year to date through December 2002." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events_archive.html) (Oct. 2009).

U.S. Energy Information Administration. (2004). "B2: Major disturbances and unusual occurrences, year to date through December 2003." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events_archive.html) (Oct. 2009).

U.S. Energy Information Administration. (2005). "B2: Major disturbances and unusual occurrences, year to date through December 2004." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events_archive.html) (Oct. 2009).

U.S. Energy Information Administration. (2006). "B2: Major disturbances and unusual occurrences, year to date through December 2005." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events_archive.html) (Oct. 2009).

U.S. Energy Information Administration. (2007). "B2: Major disturbances and unusual occurrences, year to date through December 2006." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events_archive.html) (Oct. 2009).

U.S. Energy Information Administration. (2008). "B2: Major disturbances and unusual occurrences, year to date through December 2007." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events_archive.html) (Oct. 2009).

U.S. Energy Information Administration. (2009a). "B2: Major disturbances and unusual occurrences, year to date through December 2008." (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events.html) (Oct. 2009).

U.S. Energy Information Administration. (2009b). “B1: Major disturbances and unusual occurrences, year to date through June 2009.” (http://www.eia.doe.gov/cneaf/electricity/page/disturb_events.html) (Oct. 2009).

U.S. Energy Information Administration. (2009c). “1990 – 2007 number of retail customers by state by sector (EIA-861).” (http://www.eia.doe.gov/cneaf/electricity/epa/epa_sprdshts.html) (Oct. 2009).

Whitson, J. C. and Ramirez-Marquez, J. (2009). “Resiliency as a component importance measure in network reliability.” *Rel. Eng. Syst. Safety*, 94(10), 1685-1693.

Xu, W., Hong, L., He, L., Wang, S., and Chen, X. (2011). “A supply-driven dynamic inoperability input-output price model for interdependent infrastructure systems.” *J. Infrastruct. Syst.*, in press. doi:10.1061/(ASCE)IS.1943-555X.0000058.

Zobel, C. W. (2011). “Representing perceived tradeoffs in defining disaster resilience.” *Decision Support Syst.*, 50(2), 394-403.