

# A Censored Random Coefficients Model for Pooled Survey Data With Application to the Estimation of Power Outage Costs

Klaus Moeltner<sup>1</sup>

Department of Economics, University of Washington

David F. Layton

Department of Environmental Science & Policy, University of California, Davis

## Abstract:

*In many surveys multiple observations on the dependent variable are collected from a given respondent. The resulting pooled data set is likely to be censored and to exhibit cross-sectional heterogeneity. We propose a model that addresses both issues by allowing regression coefficients to vary randomly across respondents and by using the Geweke-Hajivassiliou-Keane simulator and Halton sequences to estimate high-order probabilities. We show how this framework can be usefully applied to the estimation of power outage costs to firms using data from a recent survey conducted by a U.S. utility. Our results strongly reject the hypotheses of parameter constancy and cross-sectional homogeneity.*

*JEL classifications:* C15, C24, Q41

*Key words:* Cross-sectional heterogeneity; Multivariate censoring; Random coefficients; Simulated maximum likelihood; Power outage costs

---

<sup>1</sup> Corresponding author

Department of Economics, University of Washington

Box 353330 / Seattle, WA 98195-3330

Phone: (206) 543-5955 (work),

(206) 284-6625 (home)

Fax: (206) 685-7477

e-mail: [moeltner@u.washington.edu](mailto:moeltner@u.washington.edu)

We thank Michael J. Sullivan for his helpful comments.

## Introduction

In many panel data sets groups of observations on the dependent variable are correlated and censored at some threshold level for a given cross-sectional unit. The estimation of the resulting multivariate tobit model requires the calculation of high-dimensional probability integrals. This procedure is computationally involved beyond an order of three or four given existing software capabilities. Recent advances in the development of programming techniques to simulate high-order joint probabilities (e.g. Börsch-Supan and Hajivassiliou, 1993, Hajivassiliou et al., 1996) now allow for the consistent estimation of such models. For example, Hajivassiliou (1994) shows how these simulation techniques can be used to estimate a censored panel data model in the context of external financial crises of developing countries. Feenberg and Skinner (1994) apply one of these simulators, the Geweke-Hajivassiliou-Keane (GHK) recursive conditioning method, in a multivariate tobit estimation of a panel data set on health care expenditures.

Data sets with a panel-like structure can also result from surveys that collect multiple observations on the dependent variable from a given individual, household, or firm based on a grouping factor other than time periods. For instance, in many studies consumers are asked to report expenditures on a variety of goods. A recent example is the study by Cornick et al. (1994) on household purchases of different types of milk. In other applications, especially in the fields of labor and development economics, surveys frequently capture the allocation of time of household members across a variety of activities (e.g. Bhargava, 1997, Skoufias, 1993). In the context of estimating the value of energy reliability, commercial and industrial electricity customers are often asked to provide estimates of power outage costs associated with a set of

interruption scenarios (e.g. Beenstock et al., 1997). A fourth area of application likely to generate such data sets are contingent valuation-type studies that elicit respondents' willingness to pay (WTP) for a variety of different policy programs or quality levels. For instance, Boyle et al. (1993) administer a survey to boaters to elicit their WTP for white water rafting at different hypothetical water flow levels.

In many of these cases, the resulting pooled regression model exhibits cross-sectional heterogeneity. Thus, observations associated with a given individual or firm<sup>2</sup> are likely to be correlated. As is well known, the use of unadjusted OLS when disturbances are correlated yields inefficient parameter estimates and biased standard errors. Moulton (1986) shows that these problems are generally exacerbated when the error correlation stems from intra-unit heterogeneity.<sup>3</sup>

In addition, “zero” is often a valid answer in these studies. This introduces a censoring aspect into these models. If there are multiple zero responses per unit, the high-order probability integral problem described above in the context of panel data estimation arises. This suggests that the simulation techniques for joint probability terms, such as the GHK procedure, could be successfully employed in the estimation of such survey-generated, censored pooled regression models (CPRMs). In past studies using CPRMs, researchers have generally circumvented the computational hurdles associated with high-dimensional integrals by either assuming independence of intra-unit observations (e.g. Beenstock et al., 1997, Woo and Pupp, 1992), or by restricting the number of categories associated with each unit to a dimension that is compatible

---

<sup>2</sup> Henceforth, individual survey respondents will be alternatively referred to as “cross-sectional units”, or simply “units”.

<sup>3</sup> A related discussion in the context of CV-type studies is offered in Poe et al. (1997).

with standard computational algorithms (e.g. Cornick et al., 1994). In this paper, we specify a pooled regression model with multivariate censoring that extends existing CPRMs by allowing coefficients to vary over cross-sectional units, and by incorporating the GHK simulator and SML techniques in the estimation process. We apply this specification to a set of data on power outage cost estimates stemming from a recent survey of commercial/industrial customers conducted by a U.S. utility.

The remainder of this text is structured as follows: We start with a discussion of the econometric characteristics of the model, as well as estimation techniques. The next section briefly presents literature and data relevant to the application of our model to outage cost analysis. Estimation results and cost predictions are discussed in the fourth section. We summarize our findings in section five.

## The Model

We assume that in a given survey, multiple answers are collected from each respondent. Each answer corresponds to some category of the main grouping criterion. Such a category could be the expenditure on a specific consumer good, the number of hours allocated to some daily activity, or the WTP for a specific policy program. For simplicity, and in view of our application below, we shall call these categories “scenarios”, and assume that the answers solicited from each respondent take the form of dollar amounts. Each response to a specific scenario by a given survey participant constitutes an observation on the dependent variable. We stipulate that these observations are generated by some underlying latent variable, and that they are censored at zero. To make censoring an important feature of the model, we further assume

that there are a substantial number of “zero dollar” observations within and across cross-sectional units. We model this censoring aspect with a generic tobit specification:

$$\begin{aligned} y_{is}^* &= f(x_{is}, \mathbf{b}, \mathbf{e}_{is}) \\ y_{is} &= \max(0, y_{is}^*) \end{aligned} \tag{1}$$

Where  $y_{is}^*$  is the latent value of the dependent variable corresponding to an observation associated with unit  $i$  and scenario  $s$ ,  $y_{is}$  is the observed value,  $x_{is}$  represents a set of scenario features,  $\mathbf{b}$  is a vector of coefficients, and  $\mathbf{e}_{is}$  is a random error term.

In general, each response will depend on unit-specific characteristics, and scenario features. However, the inclusion of observed attributes of respondents in our CPRM may cause omitted variable problems, since they are likely to be correlated with unobserved error components. For example, in the context of our application below, firms’ cost estimates associated with a specific interruption scenario may well depend on firm characteristics that are not captured in a given survey, such as certain details of the production process, or the sensitivity to interruptions of existing machinery. Many of these unobserved components will, in turn, be correlated with more easily observable firm attributes such as energy consumption for a given time period or peak demand figures. Therefore, we consider only scenario features as explanatory variables in our model.

In addition, we link unit-specific observations by allowing parameter sets associated with a given respondent to vary randomly around a common mean-coefficient vector. This translates into the assumption that scenario features have a different effect on different cross-sectional units, which is intuitively sound in many applications. Our random parameter specification is similar to the model proposed by Swamy (1970). For a given unit  $i$ :

$$\begin{aligned}
y_i^* &= x_i \beta_i + e_i && \text{where } \beta_i = (\bar{\mathbf{b}} + a_i) \\
E(\mathbf{e}_i) &= 0, && E(\mathbf{e}_i \mathbf{e}_i') = s_e^2 I_T \quad i = j, \\
&&& = 0, \quad i \neq j, && (2) \\
E(a_i) &= 0 && E(a_i a_j') = ? = \begin{bmatrix} s_{11} & \cdots & s_{1k} \\ \vdots & \ddots & \vdots \\ s_{k1} & \cdots & s_{kk} \end{bmatrix} && \text{if } i = j, \\
&&& = 0, \quad i \neq j
\end{aligned}$$

$Y_i^*$  is a  $T \times 1$  vector of latent variable values associated with each observation.  $X_i$  is a  $T \times k$  matrix of scenario characteristics.<sup>4</sup> The  $k \times 1$  coefficient vector associated with each unit consists of a common mean vector  $\bar{\mathbf{b}}$ , and a vector  $a_i$  that indicates unit-specific deviations from this set of mean parameter values. We assume that these deviations are uncorrelated across units, and are distributed with mean vector zero and the common variance-covariance matrix  $\Delta$ . The diagonal and off-diagonal elements of  $\Delta$  correspond, respectively, to the variance and covariance terms of coefficients associated with scenario characteristics. The explicit estimation of these parameters is, by itself, an important objective in many studies. We will demonstrate this added benefit of a random coefficients specification in our application below.

The term  $\varepsilon_i$  is a normally distributed random vector with a mean vector of zeros and the common variance matrix  $\sigma_\varepsilon^2 I_T$  for all units. Consequently, the unit-specific coefficient vector  $\beta_i$

---

<sup>4</sup> For convenience, we assume that our data set includes the same number of rows of observations for each cross-sectional unit. The extension to a model with unbalanced cross-sectional observation sets is straightforward.

and latent variable vector  $y_i^*$  are random variables with the following multivariate normal distribution:

$$\begin{aligned} \mathbf{b}_i &\sim mvn(\bar{\mathbf{b}}, \Delta); \\ y_i^* &\sim mvn(x_i \bar{\mathbf{b}}, \Omega_i), \quad \text{where } \Omega_i = x_i \Delta x_i' + \mathbf{s}_e^2 I_T \end{aligned} \quad (3)$$

Thus, our pooled model over all cross-sectional units takes the form:

$$\begin{aligned} y^* &= X \bar{\beta} + \tilde{X} a + e \quad \text{where} \\ y^* &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad E[aa'] = S = \begin{bmatrix} \Delta & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Delta \end{bmatrix} \\ E(ee') &= s_e^2 I_{nT} \\ E(y^* y^{*'}) &= O = \tilde{X} S \tilde{X}' + s_e^2 I_{nT} \quad n = \text{number of units} \end{aligned} \quad (4)$$

Finally, like Swamy, we assume that  $x_i$ ,  $a_i$ , and  $\varepsilon_i$  are uncorrelated within and across firms. If censoring was not an issue, estimation of our model using generalized least squares (GLS) techniques would be straightforward.<sup>5</sup> Capturing the presumed strong censoring aspect prevalent in our data, however, requires the switch from least squares to a framework of joint probabilities and maximum likelihood estimation.

Given our assumption of independence of error terms across firms, the likelihood function will be a product of likelihood segments for each unit. In general, the joint probability of observing the T latent variables associated with a given firm is:

---

<sup>5</sup> See for example Swamy (1970).

$$p(y_{i1}^*, y_{i2}^*, \dots, y_{iT_i}^*) = f(y_{i1}^*, y_{i2}^*, \dots, y_{iT_i}^* | x_i \bar{\mathbf{b}}, \Omega_i) = f(v_{i1}, v_{i2}, \dots, v_{iT_i} | x_i \bar{\mathbf{b}}, \Omega_i) \quad (6)$$

where  $f(\cdot)$  denotes the multivariate normal density function, and  $\mathbf{v}$  is the combined error term vector, i.e.

$$v_i = x_i a_i + \mathbf{e}_i \quad (7)$$

If there are no “zero dollar” observations in a unit’s bundle of responses, (6) constitutes the contribution of this unit to the likelihood function. If some of the latent variable observations, say  $p$  out of  $T$ , are non-positive, we observe:

$$y_{i1} = 0, y_{i2} = 0, \dots, y_{ip} = 0, y_{i,p+1} = y_{i,p+1}^*, \dots, y_{iT} = y_{iT}^* \quad (8)$$

where the individual observations are sorted for ease of exposition. The joint probability for this outcome is:

$$\begin{aligned} & p(y_{i1} = 0, y_{i2} = 0, \dots, y_{ip} = 0, y_{i,p+1} = y_{i,p+1}^*, \dots, y_{iT} = y_{iT}^* | x_i \bar{\mathbf{b}}, \Omega_i) = \\ & p(v_{i1} \leq -\bar{\mathbf{b}}' x_{i1}, v_{i2} \leq -\bar{\mathbf{b}}' x_{i2}, \dots, v_{ip} \leq -\bar{\mathbf{b}}' x_{ip}, v_{i,p+1}, \dots, v_{iT} | x_i \bar{\mathbf{b}}, \Omega_i) = \\ & \int_{-\infty}^{-\bar{\mathbf{b}}' x_{i1}} \int_{-\infty}^{-\bar{\mathbf{b}}' x_{i2}} \dots \int_{-\infty}^{-\bar{\mathbf{b}}' x_{ip}} f(v_{i1}, v_{i2}, \dots, v_{iT} | x_i \bar{\mathbf{b}}, \Omega_i) dv_{ip} \dots dv_{i1} \end{aligned} \quad (9)$$

We denote this unit-specific segment of the likelihood function as

$$\ell_i(\bar{\mathbf{b}}, \Omega_i; y_i, x_i) \quad (10)$$

It is convenient to separate the censored from the uncensored components in (9). The theoretical underpinnings for this process are described in Pudney (1989). For an empirical application see Cornick et al. (1994). The process is based on the partitioning of  $f(\cdot)$  with respect to its censored (denoted by superscript “c”) and uncensored (denoted by superscript “uc”) segments, and applying Baye’s rule:

$$\begin{aligned} f(v_{i1}, v_{i2}, \dots, v_{iT} | x_i \bar{\mathbf{b}}, \Omega_i) &= f(v_i^c | v_i^{uc}, x_i \bar{\mathbf{b}}, \Omega_i) \cdot f(v_i^{uc} | x_i \bar{\mathbf{b}}, \Omega_i^{uc}) = \\ &= f(v_i^{uc} | x_i \bar{\mathbf{b}}, \Omega_i^{uc}) \cdot f(v_i^c | x_i \bar{\mathbf{b}}, u_i^*, \Omega_i^*) \end{aligned} \quad \text{where}$$

$$\begin{aligned} \Omega_i^{uc} &= E(v_i^{uc} v_i^{uc'}) \\ \Omega_i^* &= \Omega_i^c - \Omega_i^{c,uc} (\Omega_i^{uc})^{-1} \Omega_i^{uc,c} \\ \Omega_i^c &= E(v_i^c v_i^{c'}) \\ \Omega_i^{c,uc} &= \Omega_i^{uc,c'} = E(v_i^c v_i^{uc'}) \quad \text{and} \\ u_i^* &= \Omega_i^{c,uc} (\Omega_i^{uc})^{-1} v_i^{uc} \end{aligned} \quad (11)$$

Thus, using (9),

$$\begin{aligned} \ell_i(\bar{\mathbf{b}}, \Omega_i; y_i, x_i) &= f(v_i^{uc} | x_i \bar{\mathbf{b}}, \Omega_i^{uc}) \cdot \int_{-\infty}^{-x_i^c \bar{\mathbf{b}}} f(v_i^c | x_i \bar{\mathbf{b}}, u_i^*, \Omega_i^*) dv_i^c = \\ &= f(v_i^{uc} | x_i \bar{\mathbf{b}}, \Omega_i^{uc}) \cdot F(v_i^c | x_i \bar{\mathbf{b}}, u_i^*, \Omega_i^*) \end{aligned} \quad (12)$$

where, for notational simplicity,  $x_i^c$  refers to the whole bundle of censored observations.  $F$  is the multivariate normal cumulative distribution function (cdf). We can then specify the segment of the log-likelihood function for a cross-sectional unit as

$$\ln(\ell_i) = \ln[f(v_i^{uc} | x_i, \bar{\mathbf{b}}, \Omega_i^{uc})] + \ln[F(v_i^c | x_i, \bar{\mathbf{b}}, u_i^*, \Omega_i^*)] = \ln(\ell_i^{uc}) + \ln(\ell_i^c) \quad (13)$$

Accordingly, the log-likelihood function for the whole sample is

$$\ln(L) = \sum_{i=1}^n [\ln(\ell_i^{uc}) + \ln(\ell_i^c)] \quad (14)$$

The first term in (13) is simply a joint density, while the second term describes a joint cdf, with a potentially up to T-fold integral (if all dollar reports are zero). Since the numerical evaluation of high dimensional integrals beyond an order of three or four is impractical given currently available maximization routines, we apply simulation techniques to estimate these components.

In recent years several simulation methods for the estimation of joint probability terms have been developed by econometricians. Hajivassiliou et al. (1996) test 13 of these techniques, and find the Geweke-Hajivassiliou-Keane (GHK) simulator to provide for the overall best estimation results. As discussed in Börsch-Supan and Hajivassiliou (1993), the simulated joint probabilities generated by the GHK algorithm have the desired property of being unbiased estimates of the true probabilities, a continuous and differentiable function of the parameters of the model, and bounded by zero and one. The individual steps of the GHK procedure in the context of our model are shown in the Appendix.

The simulated joint probabilities of observing censored cost reports replace the second term in (14) for all relevant cross-sectional units. The resulting simulated log-likelihood function

$$\ln(\tilde{L}) = \sum_{i=1}^n [\ln(\tilde{\ell}_i^c) + \ln(\ell_i^{uc})] \quad (15)$$

can then be maximized with respect to the elements of  $\bar{\mathbf{b}}$  and  $\Omega_i$  using conventional MLE algorithms.

## Application

In this section we show how the model outlined above can be applied to the econometric estimation of survey-generated data on expected costs caused to commercial and industrial customers by unannounced, transmission and distribution (T&D)-type power interruptions.

In recent years electric power providers and regulators have become increasingly aware that the exclusive use of conventional engineering criteria<sup>6</sup> in designing, operating, and maintaining electricity supply systems can lead to economically inefficient investments. Traditionally, power utilities supply energy to a broad spectrum of customers, who differ in their preferences and requirements for service quality and reliability.<sup>7</sup> Efficient supply planning and

---

<sup>6</sup> Examples for such criteria are loss of load expectation, reserve margin, or failure contingencies (see Sullivan and Keane, 1995).

<sup>7</sup> As described in Woo and Pupp (1992), service quality refers to the provision of electricity within acceptable frequency and voltage ranges, while service reliability is defined as the utility's ability to deliver uninterrupted energy flows. The focus of this paper is on the latter, although many of the techniques and results presented could feasibly be applied to studies on power quality.

optimal pricing of electricity would capture these demand heterogeneities, such that, at the margin, expenditures on service improvements and resulting benefits to consumers are equal. However, the demand for energy reliability is, in general, not readily observable due to a lack of market-based mechanisms that would allow customers to signal their preferences to power suppliers.

Over the last 10 to 15 years, various theoretical and applied economic approaches to estimate the marginal value of service reliability have been suggested in the resource and energy literature.<sup>8</sup> Many of these studies use survey-based approaches in their modeling and estimation process. In most survey-based reliability studies focusing on firms, respondents are asked to provide cost estimates associated with various outage scenarios. These reports together with data on firm characteristics and outage features are then used to investigate marginal effects of firm and outage characteristics on outage costs, and to generate cost estimates for specific outage types and firm groups. Most of the existing studies on this topic use either a tabular/graphical presentation of descriptive statistics (Billington et al., 1986, Wacker et al., 1985), or ordinary least squares (OLS) regressions in their econometric estimation process (Woo et al., 1991, Analysis Group, 1990). Only a few of them take account of the fact that cost data is censored from below at zero, and adjust their model specification accordingly (e.g. Beenstock et al., 1997,<sup>9</sup> Woo and Train, 1988).

In the regression-type studies cited above, the correlation of cost reports associated with a given outage scenario is implicitly specified as zero for both intra- and inter-firm observations.

---

<sup>8</sup> For an overview and comparative evaluation of the different strategies see Caves et al. (1990, 1992).

<sup>9</sup> This citation refers to the “Subjective Evaluations” part of their study.

In other words, each row of observations is treated as independent from all others. While it can be reasonably argued that observations are uncorrelated across firms, the assumption of intra-firm independence is not likely to hold in reality. Since many firm characteristics are generally unobserved and thus not explicitly included in the estimation model, it can be expected that regression error terms are in fact correlated across firm-specific observations.<sup>10</sup>

In brief, the set of data stemming from these scenario-based outage cost surveys exhibits all the features of a CPRM described in earlier parts of this text. Also, allowing regression coefficients to vary randomly over firms is intuitively attractive in this context. It is highly likely that due to heterogeneity individual firms will be affected differently by a given outage characteristic (such as length or time of occurrence). Thus, our theoretical model outlined above should be well suited to estimate this type of data.

The utility firm in question was interested in allocating future investments in reliability equipment and services efficiently over circuits and feeder loops in its service area, both in scope and in timing. This objective required information on the sensitivity of customers to T&D type power outages in their neighborhood. The firm launched two parallel surveys, one for residential customers, the other one for commercial/industrial (C/I) units. Results from the latter were made available for this study.

In order to assure a sufficiently strong representation of C/I customers of different sizes survey designers grouped the population of firms into four categories, based on annual energy sales figures. Overall, 1451 customers were included in the survey. Firm managers were asked

---

<sup>10</sup> As discussed earlier, ignoring this correlation will reduce the efficiency of parameter estimates. This potential problem is acknowledged by Woo and Train (1988), and briefly discussed in Caves et al. (1990).

to calculate costs caused by each of six hypothetical outage scenarios. Following previous survey-based studies (e.g. Wacker et al., 1985, Woo and Train, 1988), these calculations include the value of lost production or sales, reduced staff productivity, the costs of making up for lost sales or production, equipment repair, the cost of operating backup equipment, and re-starting costs, net of any reductions in operating costs attributable to the outage.<sup>11</sup> The different outage scenarios are summarized in Table 1.

The first three columns of Table 2 show the size group labels, mean annual energy consumption, and original sample sizes for each size category. In order to ensure sufficient intra-firm variability in scenario features, only units that reported a full set of cost estimates were retained for this analysis. The last three columns in the table display the number of these remaining firms, the resulting total number of observations, and the percentage of zero cost observations in each size group.<sup>12</sup> This percentage declines as firm size increases, indicating that a larger firm is more likely to incur positive costs from a given outage than a smaller unit.

This notion extends naturally to mean and median sample statistics, as shown in Table 3. For any given scenario, median and mean estimates increase with group size. As expected, the weekday, daytime outages (excluding the momentary scenario) prompted the highest cost

---

<sup>11</sup> As discussed in Pasha, Ghaus, et al. (1989), and in Munasinghe and Sanghvi (1988), power outages can also cause indirect costs to parties that are not directly affected by the outage, but stand in some economic relation with the affected firms. For this study, only direct, short-term costs are considered.

<sup>12</sup> In each group, there are many firms with multiple (up to six) zero cost reports. Thus, there is a clear need for the application of simulation methods for high-order probability terms, as discussed earlier.

reports, ranked by outage duration. Cost estimates for a weekend, morning interruption are considerably higher than for the weekday, nighttime scenario. The 1-2 second outage generated the smallest mean and median cost estimates.

The dependent variable in our specification is expected outage cost, in dollars. The independent variables are outage duration in hours (“length”), and dummy variables for the two characteristics describing the time of occurrence of the interruption: weekday vs. weekend (“day”), and day vs. night (“time”).<sup>13</sup> We model both cost and duration in log-form.<sup>14</sup> This specification is popular in outage cost studies, since it yields an intuitively appealing concave (but never downward sloping) cost-over-duration function.<sup>15</sup> The shape of this function suggests that incremental costs are largest at the onset of an outage, but decrease as firms start to take damage control measures, e.g. by releasing employees, activating backup supplies, or by temporarily transferring business activities to unaffected locations.

We also propose an innovative way to model momentary interruptions. Clearly, intuition dictates that costs should be zero for all firms if no outage occurs. However, in the first few seconds of an interruption costs jump up instantaneously for some firms, especially in energy-

---

<sup>13</sup> We also considered firm characteristics, especially those routinely collected by the energy provider, as explanatory variables. However, as previously discussed, we found that with these characteristics we would introduce severe omitted variable and endogeneity problems into our model and therefore decided to exclude them from our estimation.

<sup>14</sup> In order to preserve zero-cost observations in log form, we follow Fische et al. (1994) by recoding original cost values from zero to one before transformation. This yields log- values of zero at the truncation point, which is convenient when introducing censoring into the model.

<sup>15</sup> See, for example, Sullivan and Keane (1995), and Woo and Pupp (1992).

intensive and high-tech industries. This introduces a conceptually awkward discontinuity at the beginning of the duration function, which makes it difficult to interpret the constant term in a regression model. We circumvent this dilemma by specifying outage costs to consist of two components, an instantaneous element and a segment associated with prolonged outage duration. We model this by including a constant term while setting the duration value for the momentary outage scenario to zero. This allows us to directly interpret the regression intercept as the expected log-cost of a momentary interruption, and its variance, as captured through the random coefficients specification of our model, as a measure of firm heterogeneity with respect to instantaneous outage costs. We assume that firms differ significantly in their sensitivity to momentary interruptions, and thus expect this variance term to be large. As will be shown below, our estimation results confirm this assumption.

We apply our model separately to each of the size groups defined above. Since the size categories are based on annual electricity consumption, this allows us to capture the effect of consumption, albeit in a discrete fashion, without exposing the model to omitted variable problems.

## Estimation Results and Cost Predictions

Table 4 summarizes the estimation results from the CPRM with random coefficients. With a few exceptions, all estimated parameters are significant at the 1% level for all size groups. As expected, all slope coefficients are positive, i.e. costs are, on average, higher on a weekday (as indicated by the dummy “day”), and during the day (as indicated by the dummy “time”), and increase with outage duration. Generally, the “time” dummy has a stronger effect on costs than the “day” term. The negative log-values of the regression intercepts for the first

three groups, and the small, positive value for group 4, respectively, indicate that the expected costs stemming from a momentary outage during the baseline time period are generally close to zero for the average firm.

The highly significant variance terms for all coefficients and size groups illustrate one of the key results of this study: Clearly, intercept and slope parameters vary considerably over cross-sectional units, indicating that a constant parameter model would be misspecified. Since the CPRM with random coefficients nests the independent tobit model, we can use a Likelihood Ratio (LR) test of the null hypothesis that all elements of  $\Delta$  are zero to compare the two specifications.<sup>16</sup> Based on the outcome of this test, we reject the null hypothesis at any reasonable level of significance for all size groups. This, in turn, implies that our assumption of correlated error terms within cross-sectional units holds.<sup>17</sup>

---

<sup>16</sup> As discussed in recent studies (e.g. Andrews, 1996), specifying parameter variances to be zero may violate the regularity conditions for maximum likelihood estimation if the likelihood function is not well defined on both sides of this zero-value threshold. For our model, we assume that  $\Omega_i$  as defined in (3) is always nonnegative-definite even for small negative values of the diagonal elements of  $\Delta$ . Thus, the likelihood function is well behaved in any neighborhood of the parameter values specified under the null hypothesis.

<sup>17</sup> The log-likelihood values for the independent tobit model are  $-7156$ ,  $-1569$ ,  $-2061$ , and  $-1147$  for sizes 1-4, respectively. The corresponding LR-statistics are  $1498.5$ ,  $276.4$ ,  $250.2$ , and  $245.96$ . The critical  $\chi^2$  value at 10 degrees of freedom and a level of significance of 1%, for example, is 23.21. Since our parameter vector is restricted under the alternative hypothesis, the LR statistic follows a mixed  $\chi^2$  distribution, and standard LR test results may be biased towards not rejecting the null hypothesis (Chen and Cosslett, 1998). However, our LR-values are well

As mentioned earlier, the explicit estimation of the variance-covariance matrix of the coefficients allows for insights beyond those provided by a model with constant parameters. For example, the relatively large variance estimates for the constant term (i.e. momentary interruptions), day, and time reflect the strong heterogeneity in expected cost reports across individual firms. This holds especially for momentary outages and for lower size groups. Thus, even though expected costs from a momentary outage are low for the whole sample, there are firms that experience considerable damage even from a very short power interruption.

Information on cross-effects of outage features can be gained by examining the off-diagonal elements of  $\Delta$ . With a few exceptions, all covariance terms are significant at a level of 5% or less.<sup>18</sup> The coefficients for day and time exhibit a positive covariance, indicating that firms that are relatively more affected by a switch from a baseline scenario interruption to a weekday, are also hit harder by a day versus a night outage. This concept extends in a similar fashion to the covariance of time and outage length, at least for the first three size groups. These results are intuitively sound. Unless a firm follows odd hours of operation, its sensitivity to power interruptions should become transparent at all scenario margins. However, as indicated by the negative and significant covariance terms involving the regression constant, this pattern is

---

above the upper bound for the critical  $\chi^2$  value for such a mixed distribution (given by  $\chi^2_{(10)}$ ), and the adjustment procedure proposed by Chen and Cosslett would not affect our test results in this case.

<sup>18</sup> We also performed a likelihood ratio test to examine the hypothesis that all *off-diagonal* terms of  $\Delta$  are jointly zero. The log-likelihood values for the resulting constrained model are -6835, -1547, -2057, and -1085 for sizes 1-4, respectively. The corresponding LR statistics are 856.4, 232.8, 242.1, and 121.4. Again, the null hypothesis is clearly rejected for all four groups.

reversed for the impact of a momentary outage versus the day, time, and length coefficients. This result suggests that a firm that experiences high instantaneous costs at the onset of a power interruption will incur relatively smaller incremental cost increases over outage duration, and for changes in the time of occurrence of the interruption. This seems reasonable for units with a high dependence on electricity and no alternative power generation capacity. Conversely, outage time and duration matters more, in terms of marginal cost increases, to firms that cope relatively well with instantaneous interruptions.

For any mix of scenario features, our model yields predictions of latent outage costs in log form. Two conceptual steps are needed to translate these estimates into actual cost forecasts. First, we convert expected latent log-costs into expected censored log-costs using the expression for the first moment of a censored normal variable<sup>19</sup>

$$E(y_s | x_s, \hat{\mathbf{b}}) = \hat{y}_s = \Phi\left(\frac{\hat{y}_s^*}{\sqrt{\hat{\Omega}_s^*}}\right) \cdot \hat{y}_s^* + \sqrt{\hat{\Omega}_s^*} \cdot \left(\frac{f(\hat{y}_s^*)}{\sqrt{\hat{\Omega}_s^*}}\right) \quad \text{where}$$

$$\hat{y}_s^* = E(y_s^* | x_s, \hat{\mathbf{b}}) = \text{expected latent cost in log - form}$$

$$\hat{\mathbf{b}} = \text{vector of parameter estimates for } \bar{\mathbf{b}} \quad (16)$$

$$\hat{\Omega}_s^* = x_s' \hat{\Delta} x_s + \hat{\mathbf{S}}_e^2 = \text{expected variance of } \hat{y}_s^*$$

$$\hat{\mathbf{S}}_e^2 = \text{estimated variance of the random error term}$$

The symbols  $\Phi$  and  $\phi$  refer to the standard normal cdf and pdf, respectively. The subscript “s” stands for a specific scenario, as described by the day, time, and length specifications in  $x_s$ .<sup>20</sup> In

<sup>19</sup> See, for example, Greene (1997), p.960.

<sup>20</sup> Since the cost prediction for a given scenario will be the same for any firm, we can omit an additional “i” subscript in this context.

the second forecasting step, censored log-costs need to be converted into absolute terms. However, actual costs follow a log-normal distribution. As is well known, simply exponentiating predicted log-costs would result in biased predictions for expected actual costs (e.g. Stynes et al., 1986). Instead, the exponent of log-costs needs to be scaled by an appropriate transformation term, i.e.

$$E\left(Y_s | x_s \hat{\mathbf{b}}\right) = t \cdot \exp(\hat{y}_s) \quad (17)$$

where  $Y_s$  stands for predicted actual costs in dollars. Several parametric and non-parametric versions of  $t$  have been suggested in the literature, mainly in the context of OLS models with homoskedastic errors.<sup>21</sup> If forecasting accuracy is important, Stynes et al. (1986) suggest to specify  $t$  as a ratio of observed over predicted values. This technique is also applied by Woo and Train (1988) in the context of their outage cost model. We take a similar approach by specifying

$$t_s = \frac{\bar{Y}_s}{\exp(\hat{y}_s)} \quad (18)$$

where  $\bar{Y}_s$  represents the sample mean associated with in-sample scenario  $s$ . Therefore, by (17) and (18), our predicted actual costs for scenarios included in the survey will be equal to the sample mean associated with these interruptions. For out-of-sample cost predictions, we

---

<sup>21</sup> E.g. Duan (1983), Goldberger (1968), and Meulenbergh (1965).

generate the necessary transformation terms by interpolation.<sup>22</sup> Table 5 summarizes the resulting cost estimates for weekday, daytime outages by size groups. As expected, outage costs increase with duration for all groups, generally at a decreasing rate. The sample means for the other two interruption types used in the survey are shown in the last two rows of the table. As indicated by the preceding discussion of parameter estimates, expected costs associated with a weekend, daytime outage are higher than those caused by a weekday, night outage for all size groups.

## Conclusion

Based on research objectives or due to cost considerations researchers often collect multiple responses associated with some categories or levels of the dependent variable from a given respondent. The resulting pooled set of data is likely to exhibit both of the following two econometric characteristics: Observations on the dependent variable are censored, and error terms include respondent-specific components and are thus correlated within each observation bundle.

In this paper we propose an econometric model that simultaneously addresses both issues. We link respondent-specific observations in an econometrically efficient and intuitively appealing way by allowing coefficient vectors associated with a given unit to deviate randomly from a common sample mean. We further demonstrate that due to recent progress made in the simulation of joint probabilities of higher order the added imposition of censoring onto our

---

<sup>22</sup> Specifically, the t-terms associated with in-sample outages display an approximately log-linear relationship to outage length. We use this fact to derive transformation terms for outages of a duration other than moment, 1,4, or 12 hours.

model does not result in insurmountable computational hurdles. Specifically, we apply the Geweke-Hajivassiliou-Keane simulator in combination with Halton sequences to evaluate multivariate cumulative distribution terms, and find that it performs well in this context.

As shown in our application, this framework allows for a more comprehensive analysis of power outage costs to commercial and industrial customers than existing models. Specifically, there are two main improvements over existing specifications stemming from our model: The number of interruption scenarios included in a given questionnaire is not limited by computational constraints associated with the calculation of high-order probabilities, and the explicit estimation of variance and covariance terms for regression coefficients allows for a better understanding of firm heterogeneity associated with each individual outage feature. Based on our results, we conclude that the assumptions of parameter constancy and independent error terms within units are clearly rejected for our sample, and probably not tenable in any outage cost estimation of data collected from heterogeneous firms.

Since our model does not include any respondent characteristics, its cost or welfare predictions are strictly scenario-specific, and unit-indiscriminant. If predictions for subgroups of respondents are desired, a model with unit-specific variables is needed. In that case, correlation of such variables with the stochastic components of the model may become a problem, and a random parameter specification may not be appropriate. However, due to the time-dependency of many respondent characteristics, the predictive power of models that include observed attributes of heterogeneous units may be short-lived. In our specification, all unit-related heterogeneity is treated as an unobserved error component. By allowing this component to be drawn from some probability distribution common to all cross-sectional units, our model is more robust to changes in individual characteristics over time than a fixed-effects specification.

Therefore, it can generate predictions associated with different scenarios that are valid for a prolonged period of time. This should be an attractive feature to many decision makers, especially when the frequent collection of data on respondent characteristics is too costly.

## Appendix A: The GHK Simulation Routine<sup>23</sup>

The following is based on Börsch-Supan and Hajivassiliou (1993) and Layton (1995):

We can express the joint probability given by the cdf term in (12) as simultaneously observing:

$$p_i = \begin{bmatrix} -\infty < v_{i1}^c - u_{i1}^* \leq -\bar{\mathbf{b}}' x_{i1} - u_{i1}^* \\ -\infty < v_{i2}^c - u_{i2}^* \leq -\bar{\mathbf{b}}' x_{i2} - u_{i2}^* \\ \vdots \\ -\infty < v_{ip}^c - u_{ip}^* \leq -\bar{\mathbf{b}}' x_{ip} - u_{ip}^* \end{bmatrix} = -\infty < v_i^c - u_i^* \leq -x_i^c \bar{\mathbf{b}} - u_i^* \quad (\text{A1})$$

We then define:

$$\Omega_i^* = \Gamma_i \Gamma_i' \quad (\text{A2})$$

where  $\Gamma_i$  is the lower triangular matrix in a Choleski decomposition.  
 $e_i$  = set of  $p$  i.i.d. standard normal random variables.

Then :

$$E(\Gamma_i e_i) = \Gamma_i \cdot E(e_i) = 0, \quad \text{and} \quad (\text{A3})$$

$$E(\Gamma_i e_i (\Gamma_i e_i)') = E(\Gamma_i e_i e_i' \Gamma_i') = \Gamma_i E(e_i e_i') \Gamma_i' = \Gamma_i I_p \Gamma_i' = \Gamma_i \Gamma_i' = \Omega_i^*$$

---

<sup>23</sup> The GAUSS code for this procedure was kindly made available by Vassilis Hajivassiliou at his London School of Economics internet site. The Matlab version of this routine, as well as programming code for the general estimation process are available from the authors upon request.

In other words,  $(v_i^c - u_i^*)$  and  $G_i e_i$  have the same multivariate normal distribution. Therefore, we can re-write (A1) as:

$$p_i = \begin{bmatrix} -\infty < l_{11}e_1 \leq -\bar{\mathbf{b}}'x_{i1} - u_{i1}^* \\ -\infty < l_{21}e_1 + l_{22}e_2 \leq -\bar{\mathbf{b}}'x_{i2} - u_{i2}^* \\ \vdots \\ -\infty < l_{p1}e_1 + l_{p2}e_2 + \dots + l_{pp}e_p \leq -\bar{\mathbf{b}}'x_{ip} - u_{ip}^* \end{bmatrix} \quad (\text{A4})$$

where  $l_{ij}$  is the  $i^{\text{th}}$  row,  $j^{\text{th}}$  column element of  $\Gamma$ , and  $e_1$  to  $e_p$  are the components of the  $e_i$  vector.

As a first step in the simulation process, we draw  $e_1$  such that:

$$-\infty < e_1 \leq (-\bar{\mathbf{b}}'x_{i1} - u_{i1}^*)/l_{11} \quad (\text{A5})$$

This can be accomplished using the integral transform theorem (see for example Greene, 1997):

$$e_1 = \Phi^{-1}\{\Phi((- \bar{\mathbf{b}}'x_{i1} - u_{i1}^*)/l_{11}) \cdot z_1\} \quad (\text{A6})$$

where  $z_1$  is random uniform  $[0,1]$ , and  $\Phi$  is the standard normal cdf. For our analysis, we used Halton sequences (Halton, 1960) to generate the  $z$ -terms. This process yielded more efficient probability estimates than uniformly random draws from the unit interval or antithetic techniques. Appendix B provides an outline of this procedure. In passing, we note that

$$p_1 = p(-\infty < e_1 \leq (-\bar{\mathbf{b}}'x_{i1} - u_{i1}^*)/l_{11}) = \Phi((- \bar{\mathbf{b}}'x_{i1} - u_{i1}^*)/l_{11}) \quad (\text{A7})$$

Given the shape of  $\Gamma_{ie_i}$  shown above, the process is recursive, i.e.  $e_1$  will figure in the drawing of  $e_2$ , both enter the drawing of  $e_3$ , and so on. This process is repeated until its final step, the calculation of  $p_p$ , given  $e_1$  through  $e_{p-1}$ :

$$p_p = P\left(-\infty < e_p \leq \left(-\bar{\mathbf{b}}'x_{i1} - u_{i1}^* - \sum_{j=1}^{p-1} l_{pj}e_j\right)/l_{pp}\right) = \Phi\left(\left(-\bar{\mathbf{b}}'x_{i1} - u_{i1}^* - \sum_{j=1}^{p-1} l_{pj}e_j\right)/l_{pp}\right) \quad (\text{A8})$$

Thus, the joint probability for our desired set of outcomes is:

$$p_i = \prod_{j=1}^p p_j \quad (\text{A10})$$

This simulation process is repeated  $R$  times,<sup>24</sup> yielding a mean joint probability of

$$\frac{1}{R} \left( \sum_{r=1}^R \prod_{j=1}^p p_{rj} \right) \quad (\text{A11})$$

This term is precisely the simulated proxy needed for our censored likelihood segment  $l_i^c$  used in (15), i.e.

$$\tilde{\ell}_i^c(\bar{\mathbf{b}}, \Omega_i^* | y_i, x_i, R) = \frac{1}{R} \left( \sum_{r=1}^R \prod_{j=1}^p p_{rj} \right) \quad (\text{A12})$$

---

<sup>24</sup> We specify  $R = 1000$ .

## Appendix B: Halton Sequences

In recent studies Train (1999) and Feenberg and Skinner (1994) discuss the merits of Halton sequences in the context of simulating high-order probabilities. Specifically, they find that Halton sequences provide for smaller standard errors of the simulator and a faster convergence rate of the associated maximum likelihood algorithm than standard uniform draws or antithetic variance reduction methods. Our own Monte Carlo tests for GHK simulators based on different drawing methods of  $\mathbf{z}$  in (A6) using original test specifications given in Börsch-Supan and Hajivassiliou (1993) confirmed these findings. The following brief outline of Halton sequences follows closely the studies by Train and Feenberg and Skinner cited above.<sup>25</sup>

Halton sequences are designed to span most efficiently the unit interval  $[0,1]$ . As such, they are often labeled as “pseudorandom” drawings. Each sequence is calculated based on a different prime number. For prime number  $p$ , for example, the first step in the Halton procedure is to divide the unit interval into  $p$  parts. The dividing points become the first elements in the sequence. Each of the  $p$  sub-divisions of the unit interval is, in turn, divided into  $p$  parts. The new dividing points constitute the next elements of the sequence, and so forth. Since early elements in a given sequence tend to be correlated over sequences with different primes, the first 10 or so elements of each sequence are usually discarded.

For our GHK application we require a matrix of Halton terms with  $i$  rows and  $r$  columns, where  $i$  is the dimension of the multivariate probability term to be simulated (i.e. between 1 and 6), and  $r$  equals the desired number of GHK repetitions (i.e. 1000). For example, the estimation

---

<sup>25</sup> The GAUSS code for generating Halton sequences is kindly provided in the appendix of the Feenberg and Skinner study. The Matlab version is available from the authors of this study.

of the likelihood contribution of a firm with 4 zero-cost observations requires the generation of four Halton sequences, based on the prime numbers 3, 5, 7, and 11, respectively.

## References

- Analysis Group, Inc. "Industrial Outage Cost Survey: Final Report," Submitted to Niagara Mohawk Power Corporation (1990).
- Andrews, Donald W. K., "Admissibility of the Likelihood Ratio Test when the Parameter Space is Restricted under the Alternative," *Econometrica* 64, (1996), 705-718.
- Beenstock, Michael, Ephraim Goldin, and Yoel Haitovsky, "The Cost of Power Outages in the Business and Public Sectors in Israel: Revealed Preferences vs. Subjective Valuation," *The Energy Journal* 18, (1997), 39-61.
- Bhargava, Alok, "Nutritional Status and the Allocation of Time in Rwandese Households," *Journal of Econometrics* 77, (1997), 277-295.
- Billington, R., G. Wacker, and R.K. Subramaniam, "Factors Affecting the Development of a Commercial Customer Damage Function," *IEEE Transactions on Power Systems* PWRS-1, (1986), 28-33.
- Börsch-Supan, Axel, and Vassilis A. Hajivassiliou, "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics* 58, (1993), 347-368.
- Boyle, Kevin J., Michael P. Welsh, and Richard C. Bishop, "The Role of Question Order and Respondent Experience in Contingent-Valuation Studies," *Journal of Environmental Economics and Management* 25, (1993), S-80-S-99.
- Caves, Douglas W., Joseph A. Herriges, and Robert J. Windle, "The Cost of Electric Power Interruptions in the Industrial Sector: Estimates Derived from Interruptible Service Programs," *Land Economics* 68, (1992), 49-61.

- Caves, Douglas W., Joseph A. Herriges, and Robert J. Windle, "Customer Demand for Service Reliability in the Electric Power Industry: A Synthesis of the Outage Cost Literature," *Bulletin of Economic Research* 42, (1990), 79-119.
- Chen, Heng Z., and Stephen R. Cosslett, "Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach," *American Journal of Agricultural Economics*, (1998), 512-520.
- Cornick, Jorge, Thomas L. Cox, and Brian W. Gould, "Fluid Milk Purchases: A Multivariate Tobit Analysis," *American Journal of Agricultural Economics* 76, (1994), 74-82.
- Duan, Naihua, "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association* 78, (1983), 605-610.
- Feenberg, Daniel, and Jonathan Skinner, "The Risk and Duration of Catastrophic Health Care Expenditures," *The Review of Economics and Statistics* 76, (1994), 633-647.
- Fishe, R.P.H., G.S. Maddala, and R.P. Trost. "Estimation of a Heteroskedastic Tobit Model." in G.S. Maddala (ed.) *Econometric Methods and Applications* (Aldershot, U.K.:Elgar, 1994, 112-122).
- Goldberger, Arthur S., "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica* 35, (1968), 464-472.
- Greene, William H. *Econometric Analysis* (Prentice Hall, 1997).
- Hajivassiliou, Vassilis, Daniel McFadden, and Paul Ruud, "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," *Journal of Econometrics* 72, (1996), 85-134.
- Hajivassiliou, Vassilis A., "A Simulation Estimation Analysis of the External Debt Crises of Developing Countries," *Journal of Applied Econometrics* 9, (1994), 109-131.

- Halton, J., "On the Efficiency of Certain Quasirandom Sequences of Points in Evaluating Multidimensional Integrals," *Numerical Mathematics*, (1960), 84-90.
- Layton, David Frost. "Specifying and Testing Econometric Models for Stated Preference Surveys," Doctoral Dissertation, Department of Economics, University of Washington, 1995.
- Meulenberg, M. T. G., "On the Estimation of an Exponential Function," *Econometrica* 33, (1965), 863-868.
- Moulton, Brent R., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics* 32, (1986), 385-97.
- Munasinghe, Mohan, and Arun P. Sanghvi, "Reliability of Electricity Supply, Outage Costs and Value of Service," *The Energy Journal* 9, (1988), 1-18.
- Pasha, Hafiz A., Aisha Ghaus, and Salman Malik, "The Economic Cost of Power Outages in the Industrial Sector of Pakistan," *Energy Economics*, (1989), 301-318.
- Poe, Gregory L., Michael P. Welsh, and Patricia A. Champ, "Measuring the Difference in Mean Willingness to Pay When Dichotomous Choice Contingent Valuation Responses Are Not Independent," *Land Economics* 73, (1997), 255-267.
- Pudney, Stephen. *Modelling Individual Choice - The Econometrics of Corners, Kinks and Holes* (Basil Blackwell, 1989).
- Skoufias, Emmanuel, "Labor Market Opportunities and Intrafamily Time Allocation in Rural Households in South East Asia," *Journal of Development Economics* 40, (1993), 277-310.
- Stynes, Daniel J., George L. Peterson, and Donald H. Rosenthal, "Log Transformation Bias in Estimating Travel Cost Models," *Land Economics* 62, (1986), 94-103.

- Sullivan, M. J. and D. M. Keane, *Outage cost estimation guide book*, in: Electric Power Research Institute, Report No. TR-106082 (San Francisco, CA, 1995).
- Swamy, P. A. V. B., "Efficient Inference in a Random Coefficient Regression Model," *Econometrica* 38, (1970), 311-323.
- Train, Kenneth, "Halton Sequences for Mixed Logit," Working Paper, Department of Economics, University of California, Berkeley (1999).
- Wacker, G., R.K. Subramaniam, and R. Billington, "Using Cost of Electric Service Interruption Surveys in the Determination of a Composite Customer Damage Function," *International Journal of Energy Systems* 5, (1985), 100-104.
- Woo, Chi-Keung, and Roger L. Pupp, "Costs of Service Disruptions to Electricity Consumers," *Energy* 17, (1992), 109-126.
- Woo, Chi-Keung, Roger L. Pupp, Theresa Flaim, and Robert Mango, "How Much Do Electric Customers Want to Pay for Reliability? New Evidence on an Old Controversy," *Energy Systems and Policy* 15, (1991), 145-159.
- Woo, Chi-Keung, and Kenneth Train, "The Cost of Electric Power Interruptions to Commercial Firms," *The Energy Journal* 9, (1988), 161-172.

Tables:

Table 1: Outage Scenarios

<b>Scenario</b>	<b>Duration (hrs)</b>	<b>Day</b>	<b>Time</b>
1	1	Weekday	10:00 AM
2	Moment (1-2 seconds)	Weekday	10:00 AM
3	4	Weekday	10:00 AM
4	12	Weekday	10:00 AM
5	1	Weekday	Midnight
6	1	Weekend	10:00 AM

Table 2: Sample Characteristics by Size Group

<b>Group</b>	<b>Group Mean Annual Energy Consumption (MWH)</b>	<b>No. of Firms (original)</b>	<b>No. of Firms (retained)</b>	<b>No. of Observations (retained)</b>	<b>Percentage of Zero Cost Observations</b>
1	12	933	671	4026	43%
2	131	189	132	792	29%
3	670	209	161	966	20%
4	3102	120	82	492	14%

Table 3: Sample Statistics for Median and Mean Cost Estimates

	Scenario											
	1		2		3		4		5		6	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean
<b>Size</b>												
1	\$135	\$540	0	\$44	432	\$1,355	900	\$2,553	0	\$69	0	\$178
2	\$400	\$1,280	0	\$149	1000	\$2,901	2650	\$8,739	0	\$323	162	\$552
3	\$1,000	\$3,368	0	\$915	4000	\$8,272	7500	\$18,501	150	\$1,470	400	\$2,086
4	\$3,000	\$8,078	100	\$861	7775	\$24,979	20000	\$57,146	500	\$3,498	900	\$4,802

Table 4: Parameter Estimates From the Censored Random Coefficients Model

Parameter	Size 1		Size 2		Size 3		Size 4	
	Estimate	s.e	Estimate	s.e	Estimate	s.e.	Estimate	s.e.
Constant	-12.861 <sup>a</sup>	0.650	-7.188 <sup>a</sup>	0.873	-2.860 <sup>a</sup>	0.520	0.200	0.678
Day	3.231 <sup>a</sup>	0.213	1.425 <sup>a</sup>	0.276	1.234 <sup>a</sup>	0.244	1.466 <sup>a</sup>	0.398
Time	7.973 <sup>a</sup>	0.515	5.886 <sup>a</sup>	0.636	2.770 <sup>a</sup>	0.273	1.224 <sup>a</sup>	0.318
Length	1.190 <sup>a</sup>	0.028	1.179 <sup>a</sup>	0.049	1.206 <sup>a</sup>	0.054	0.998 <sup>a</sup>	0.079
Var (const.)	111.543 <sup>a</sup>	11.420	73.937 <sup>a</sup>	14.158	40.361 <sup>a</sup>	5.366	34.094 <sup>a</sup>	6.364
Var (day)	17.908 <sup>a</sup>	1.442	8.421 <sup>a</sup>	1.537	7.621 <sup>a</sup>	1.257	10.965 <sup>a</sup>	2.195
Var (time)	31.916 <sup>a</sup>	4.592	28.334 <sup>a</sup>	6.526	10.006 <sup>a</sup>	1.550	6.603 <sup>a</sup>	1.355
Var (length)	0.220 <sup>a</sup>	0.021	0.202 <sup>a</sup>	0.034	0.342 <sup>a</sup>	0.054	0.415 <sup>a</sup>	0.081
Cov (const., day)	-30.599 <sup>a</sup>	2.362	-16.425 <sup>a</sup>	2.844	-8.620 <sup>a</sup>	1.850	-11.951 <sup>a</sup>	2.553
Cov (const., time)	-54.282 <sup>a</sup>	7.236	-40.693 <sup>a</sup>	9.475	-15.205 <sup>a</sup>	2.490	-8.080 <sup>a</sup>	2.386
Cov (const., length)	-2.111 <sup>a</sup>	0.317	-2.069 <sup>a</sup>	0.468	-2.200 <sup>a</sup>	0.447	-1.865 <sup>a</sup>	0.537
Cov (day, time)	13.015 <sup>a</sup>	1.783	8.209 <sup>a</sup>	2.405	2.684 <sup>a</sup>	0.934	4.278 <sup>a</sup>	1.262
Cov (day, length)	-0.045	0.112	-0.060	0.147	-0.317	0.171	-0.280	0.282
Cov (time, length)	0.887 <sup>a</sup>	0.162	0.797 <sup>a</sup>	0.289	0.393 <sup>b</sup>	0.193	-0.187	0.235
Var (eps.)	1.630 <sup>a</sup>	0.075	1.440 <sup>a</sup>	0.135	1.745 <sup>a</sup>	0.154	1.109 <sup>a</sup>	0.134
Sample size	4026		792		966		492	
Log likelihood	-6406.806		-1430.663		-1935.954		-1024.315	
Ghk replications	1000		1000		1000		1000	

<sup>a</sup>Significant at the 1 percent level    <sup>b</sup>Significant at the 5 percent level

Table 5: Predicted Average Outage Costs

Outage Type	Costs (\$)			
	Size 1	Size 2	Size 3	Size 4
<i>Weekday, daytime</i>				
Length (hrs)				
Moment <sup>s</sup>	44	149	915	861
1 <sup>s</sup>	540	1280	3368	8078
2	907	2183	5989	14765
3	1174	2693	7556	20357
4 <sup>s</sup>	1355	2901	8272	24979
5	1607	3682	10145	30173
6	1825	4455	11844	34995
7	2013	5216	13366	39466
8	2171	5961	14714	43603
9	2302	6687	15892	47424
10	2408	7393	16910	50944
11	2491	8077	17776	54179
12 <sup>s</sup>	2553	8739	18501	57146
<i>Weekday, night (1 hr)<sup>s</sup></i>	69	323	1470	3498
<i>Weekend, daytime (1 hr)<sup>s</sup></i>	178	552	2086	4802

<sup>s</sup>Sample statistic