

ESTIMATION AND TESTING FOR FUNCTIONAL FORM IN PURE PREMIUM REGRESSION MODELS

BY SCOTT E. HARRINGTON

University of Pennsylvania

ABSTRACT

Estimation of pure premiums for alternative rate classes using regression methods requires the choice of a functional form for the statistical model. Common choices include linear and log-linear models. This paper considers maximum likelihood estimation and testing for functional form using the power transformation suggested by Box and Cox. The linear and log-linear models are special cases of this transformation. Application of the procedure is illustrated using auto insurance claims data from the state of Massachusetts and from the United Kingdom. The predictive accuracy of the method compares favorably to that for the linear and log-linear models for both data sets.

KEYWORDS

Pure premium regression models, functional form, maximum likelihood estimation, power transformation.

1. INTRODUCTION

An important issue that arises in modeling claim frequency, claim severity, or pure premiums is the choice of an appropriate functional form for the statistical model. The importance of this choice is illustrated in the classical work of BAILEY and SIMON (1960). This study compared the traditional multiplicative ratemaking method used in the United States and Canada with multiplicative and additive models estimated using the minimum chi-square method.* The results of analyzing Canadian automobile insurance loss ratios for driver and merit rating classes indicated that the additive model provided a better fit than the multiplicative procedures. The multiplicative methods also appeared to produce systematic overestimates for the highest risk merit rating and driver classes.

The tendency of the traditional multiplicative method used in Canada to overestimate pure premiums for high-risk drivers was later debated by HOLMES (1970) and WILCKEN (1971). More recently, CHANG and FAIRLEY (1978, 1979) documented this tendency using data on average claims per exposure by driver class and territory for the state of Massachusetts. They also found that an additive model estimated using weighted least squares (with the square root of cell

* The minimum chi-square method for estimating a multiplicative model had earlier been considered by ALMER (1957). For a description of these and similar models and estimation techniques, see WEISBERG and TOMBERLIN (1982) and EEGHEN, GREUP and NIJSSEN (1983)

exposures as weights) did not produce overestimates for the highest risk classes and that the additive model fit the data better than either the traditional multiplicative method or a log-linear regression model. Similar results were obtained by FAIRLEY, TOMBERLIN and WEISBERG (1981) using auto insurance claims data for the state of New Jersey. Each of these studies emphasized the undesirability of using methods that overestimate pure premiums for the highest rate classes in view of the problem of automobile insurance affordability.

SANT (1980) analyzed some of the Massachusetts data employed by Chang and Fairley using a multiplicative model with additive errors. Weighted least squares estimates of this model produced smaller estimates for the highest risk classes than the traditional multiplicative method, but the additive least squares model still provided the best overall fit. Sant suggested that the latter result could be caused by the omission of interaction terms in the multiplicative model rather than an incorrect functional form. SAMSON and THOMAS (1984) focused on the influence of interaction terms in estimating additive and log-linear models of pure premiums using data for a United Kingdom motor insurance account. In contrast to the previous studies, the log-linear model fit the data much better than the additive model. However, an additive model with interaction terms selected by stepwise regression provided a better fit than either the additive or log-linear models without interaction terms.

A number of authors have considered other alternatives to purely additive or multiplicative models. BAILEY and SIMON (1960) estimated a model with both additive and multiplicative terms. CHAMBERLAIN (1980) fitted a multiplicative model to the residuals from estimating an additive model. Both methods generally failed to improve upon both the additive and multiplicative models in terms of goodness of fit. WEISBERG, TOMBERLIN, and CHATTERJEE (1984) analyzed a pure premium model with both additive and multiplicative terms. They also considered models that involved separate estimation of frequency and severity. Comparisons with purely additive and multiplicative models using Massachusetts and New Jersey data suggested that the functional form for the pure premium models and separate estimation of frequency and severity had relatively little impact on predictive accuracy.*

FREIFELDER (1984) predicted the patterns of estimation errors that would be expected if the true model were additive and a multiplicative model were to be assumed in estimation and vice versa. Using the Massachusetts data employed by Chang and Fairley, he conducted chi-square tests to determine whether the error patterns for the methods were consistent with those predicted. The results suggested that both the traditional multiplicative method and the additive least squares model could be rejected for collision coverage and that the multiplicative method, but not the additive model, could be rejected for combined compulsory

* WEISBERG, TOMBERLIN and CHATTERJEE (1984) also examined both the simple use of individual cell pure premiums and the use of Bayes methods that combined individual cell pure premiums with model estimates. The use of cell pure premiums performed well in terms of predictive accuracy for the Massachusetts data, which had very large cells. The Bayes estimates generally outperformed both the model-based estimates and the use of cell pure premiums.

coverages. The study is important because it relies on statistical testing in comparing functional forms. A limitation of the test procedure is that it only considers the signs of the errors and not their magnitude (cf. SAMSON and THOMAS (1984)). Furthermore, an alternative model is not implied when both the additive and multiplicative forms are rejected by this test.

The choice between additive and multiplicative models of automobile insurance claims appears to have been of less concern to European researchers than it has in North America. Most European researchers also have focused on models of claim frequency and severity rather than pure premium models.* Despite these differences in emphasis, the choice of functional form is likely to be important in all types of claims modeling (see, for example, ALBRECHT (1983)).

This paper considers estimation and testing for functional form in pure premium regression models. The procedure suggested by BOX and COX (1964) and extended by LAHIRI and EGY (1981) for the case of heteroscedastic disturbances is applied to Massachusetts automobile insurance claims data used by CHANG and FAIRLEY (1979) and to the United Kingdom data used by SAMSON and THOMAS (1984). The procedure allows for a continuum of functional forms based on the parameter of a power transformation. The additive (linear) and log-linear functional forms are special cases of the transformation. Maximum likelihood estimates of this parameter are obtained under the normality assumption. Likelihood ratio tests are used to test whether the linear and log-linear functional forms are consistent with the data. The normality assumption underlying the estimation method and hypothesis tests is examined, and the predictive accuracy of the alternative functional forms is compared.

The study follows almost all previous work that has focused on the choice between additive and multiplicative models in that it deals exclusively with estimation and testing for functional form in pure premium models. The flexible functional form approach also could be applied to separate estimation of frequency and severity using a variety of distributional assumptions. Extensions in this direction would be desirable, since the relatively simple functional forms examined in this paper may only roughly approximate those that might be implied by explicit modeling of frequency and severity.

Section 2 provides details concerning the models and estimation methods. Criteria for comparing performance of the models are described in Section 3. The illustrative results for the Massachusetts and United Kingdom data are presented in Section 4. Section 5 contains suggestions for further study.

2. MODELS AND ESTIMATION METHOD

Subject to degrees of freedom considerations, the methodology illustrated in this paper could be employed for a pure premium model with any number of factors, levels, interaction terms, or continuous regressors. For simplicity, the models first are described assuming that there are two factors, 1 and 2, with an unspecified

* COUTTS (1984) summarizes much of this work. Also see EEGHEN, GREUP and NIJSSEN (1983)

number of levels and no interaction terms. The additive or (using the more common regression terminology) linear model may be expressed as:

$$(2.1) \quad \bar{p}_{ij} = b_{1i} + b_{2j} + e_{ij}$$

where \bar{p}_{ij} is total claims for cell ij divided by n_{ij} , the number of exposures in the cell, b_{1i} is the effect of the i th level of factor 1, b_{2j} is the effect of the j th level of factor 2, and e_{ij} is a disturbance with zero mean.

The variance of e_{ij} is assumed to equal $s^2(n_{ij})^{-d}$ where s^2 is constant across cells. For most of the empirical work, d is constrained to equal one. In this case, equation (2.1) corresponds to the additive model employed by CHANG and FAIRLEY (1978, 1979) and others. In part of the analysis, this constraint is dropped and d is treated as an unknown parameter to be estimated. A variety of other assumptions concerning disturbance variance could be made. For example, the variance could be related to the moments of an assumed underlying distribution for claim frequency and, perhaps, severity (see BAILEY and SIMON (1960), SEAL (1968), WEISBERG and TOMBERLIN (1982), and ALBRECHT (1983)).

Whether refinements of this type would be likely substantively to improve predictive accuracy is an empirical question. In preliminary work, the linear model was estimated assuming that disturbance variance was equal to $k(b_{1i} + b_{2j})/n_{ij}$ where k is a constant. As shown by WEISBERG and TOMBERLIN (1982), this assumption would be appropriate if claim frequency was Poisson distributed and the distribution of claim severity was identical for each cell. A two-pass weighted least squares regression procedure was employed with the first-pass estimates of b_{1i} and b_{2j} used to construct the weights for the second pass. Iteration until convergence would have been possible (cf. SEAL (1968)), but it would not change the estimator's asymptotic properties (see AMEMIYA (1973)). The parameter estimates and predicted pure premiums for this method were very similar to those obtained using weighted least squares with the square root of exposures as weights.

The log-linear model may be expressed as:

$$(2.2) \quad \log(\bar{p}_{ij}) = b_{1i} + b_{2j} + e_{ij}$$

where the same assumptions are made concerning e_{ij} as for the linear model. The log-linear model has been used by CHANG and FAIRLEY (1978, 1979), FAIRLEY, TOMBERLIN and WEISBERG (1981), and SAMSON and THOMAS (1984).* It may be expected to provide pure premium estimates that are similar to those that would be obtained using either the traditional multiplicative method or the

* As was pointed out to the author by H Buhlmann and A Gisler, the b s in model (2.2) (and model (2.3) described below) might be expected to depend on the number of exposures in the given cell. The treatment of these parameters as constants for cells of different sizes in this study is reasonable in view of the large number of exposures in each cell in the data used. In general, a theoretically preferable procedure for estimating models using transformed data would be to use the mean of the transformed values of the individual data as opposed to using the transformed value of the mean. This observation highlights the approximate nature of model (2.2) (and model (2.3) when $c=0$). The log-linear functional form could apply to the expected pure premium for an individual exposure. It could not apply to actual claims for an individual exposure if zero claims were possible.

multiplicative least squares model employed by SANT (1980). Estimation of the log-linear model is considerably simpler than multiplicative least squares, and the statistical properties of the estimates under the normality assumption are well-known, in contrast to those for the traditional multiplicative method.

The BOX and COX (1964) procedure, as extended by LAHIRI and EGY (1981) to allow for heteroscedasticity, posits that for some value of the parameter c , $\bar{p}_y^{(c)}$ will be linear in b_{1i} and b_{2j} , i.e.:

$$(2.3) \quad \bar{p}_y^{(c)} = b_{1i} + b_{2j} + e_{ij}$$

where

$$(2.4) \quad \bar{p}_y^{(c)} = (\bar{p}_y^c - 1)/c.$$

When $c = 1$, model (2.3) is equivalent to the linear model. Since the limit of $\bar{p}_y^{(c)}$ as c goes to zero is $\log(\bar{p}_y)$, it is equivalent to the log-linear model when $c = 0$.

To facilitate explanation of the estimation procedures used, models (2.1), (2.2), and (2.3) may be described as:

$$(2.5) \quad y_{ij} = b_{1i} + b_{2j} + \sum_{i \neq 1} (b_{1i} - b_{11})D_{1i} + \sum_{j \neq 1} (b_{2j} - b_{21})D_{2j} + e_{ij}$$

where y_{ij} equals \bar{p}_y , $\log(\bar{p}_y)$, or $\bar{p}_y^{(c)}$, D_{1i} is a dummy variable that equals one for the i th level of factor 1 and zero for all other levels, and D_{2j} is a dummy variable that equals one for the j th level of factor 2 and zero for all other levels. Since (2.5) is linear in $b_{1i} + b_{2j}$ and the coefficients on each dummy variable, it can be estimated using standard linear methods. Estimation of (2.5) will not provide estimates of each b_{1i} and b_{2j} , but it will yield estimates of $b_{1i} + b_{2j}$ for each cell. These estimates can be used to predict the mean pure premium for each cell.*

The linear and log-linear versions of (2.5) may be estimated directly using least squares. If the e_{ij} s are assumed to be normally distributed with mean zero and variance $s^2(n_{ij})^{-d}$, the concentrated log-likelihood function for the Box-Cox model is given by (see LAHIRI and EGY (1981)):

$$(2.6) \quad L = K - (d/2) \sum \log(1/n_{ij}) + (c-1) \sum \log(\bar{p}_y) - (N/2) \log(\hat{s}^2)$$

where K is a constant, N is the total number of cells, \hat{s}^2 is $1/N$ times the residual sum of squares from the application of weighted least squares to equation (2.5) using $n_{ij}^{d/2}$ as the weight for cell ij , and the summation is over all cells. Hence, the coefficients in (2.5) influence L through their impact on \hat{s}^2 .

If it is assumed that d equals one, maximum likelihood estimates of the coefficients and c may be obtained by searching within some reasonable range for the value of c that maximizes L . The procedure used in this study was to

* The omission of one dummy variable for one level of each factor in (2.5) makes the resultant design matrix nonsingular. The choice of the level that is omitted is arbitrary and does not affect the predicted values.

specify a range for c and calculate the value of L at 0.05 increments for c within this range. If the maximum value occurred at an endpoint, the range was modified until an internal maximum was obtained. When d is allowed to vary, maximum likelihood estimates of the coefficients in (2.5), c , and d may be obtained by searching for values of c and d that maximize L . The zig-zag iterative search technique due to OBERHOFER and KMENTA (1974) and employed by Lahiri and Egi was used in this case.

3. COMPARISON CRITERIA

BAILEY and SIMON (1960) suggested four criteria that rate relativities or, more generally, a pure premium estimation procedure should meet that have commonly guided later researchers in selecting between competing methods. These criteria include (1) credibility, (2) minimal departure from the raw data (goodness of fit), (3) the prediction error for each subgroup can reasonably be attributed to chance (random prediction errors), and (4) the predictions should be balanced for each major class and in total.

The estimation technique for the linear and log-linear models involves minimizing the sum of squared errors using the data weighted by the square root of the number of exposures in each cell. It thus reflects the relative credibility of experience in each cell. The log-likelihood that is maximized for the Box and Cox procedure also reflects the number of exposures across cells.

With regard to goodness of fit, an ideal procedure might be to analyze forecast errors for a period following the estimation period. While this has been done in some studies (e.g., JOHNSON and HEY (1971), COUTTS (1984), and WEISBERG, TOMBERLIN and CHATTERJEE (1984)), the more common procedure of comparing actual and predicted cell means during the estimation period is used in this paper. Two measures of overall fit are employed:

$$\text{Mean Squared Error} = (1/N) \sum n_y (\bar{p}_y - \hat{p}_y)^2$$

$$\text{Mean Absolute Error} = (1/N) \sum n_y |\bar{p}_y - \hat{p}_y|$$

where \hat{p}_y is the predicted value of p_y , the true pure premium.

The predicted values are calculated from:

$$\text{Linear: } \hat{p}_y = f$$

$$\text{Log-Linear: } \hat{p}_y = \exp(f + \hat{s}^2 / (2n_y))$$

$$\text{Box and Cox: } \hat{p}_y = (cf + 1)^{1/c}, \quad c \neq 0$$

where f equals the predicted value of \bar{p}_y , $\log(\bar{p}_y)$, and $\bar{p}_y^{(c)}$ using the weighted least squares estimates of the coefficients in (2.5) for the linear, log-linear, and Box and Cox models, respectively. The predictor for the linear model will be unbiased and that for the log-linear model will be consistent if the assumptions underlying either model are satisfied. The predictor for the Box and Cox model

generally will not be consistent (unless $c = 0$, in which case the log-linear predictor is used).^{*} As discussed by NELSON and GRANGER (1979), there is no closed form for the optimal, consistent predictor for general c . While it could be calculated numerically, this was not done. Simulation results described by NELSON and GRANGER (1979) suggest that using numerical methods to obtain consistent predictions with the Box and Cox procedure may have little impact on predictive accuracy if the assumptions underlying the model are satisfied.

The principal motivation for considering the Box and Cox estimator is that by allowing the functional form to be determined by the data, it may produce predicted values with smaller mean absolute or mean squared errors than the linear and log-linear models. Its forecasting performance for economic time series has been decidedly mixed, as is illustrated by the results of NELSON and GRANGER (1979). However, these authors suggest that its weak performance may be due to the extreme nonnormality of the economic time series that they analyzed. The normality assumption may be more appropriate (i.e., violated to a lesser degree) for cross-sectional models of automobile insurance pure premiums.

If the normality assumption and certain regularity conditions were to be satisfied (see LAHIRI and EGY (1979)), a likelihood ratio test could be used to test whether the null hypothesis that either $c = 0$ or $c = 1$ is consistent with the data. For example, letting \hat{c}_{\max} denote the value of c that maximizes L , the statistic $2[L(c = \hat{c}_{\max}, d = 1) - L(c = 0, d = 1)]$, which has a limiting chi-square distribution with one degree of freedom, provides a test of the null hypothesis that $c = 0$ against the unconstrained alternative, given that $d = 1$. Analogous tests could be conducted for the null hypothesis that $c = 1$ and to test hypotheses concerning the magnitude of d .

To provide evidence of whether the normality assumption is reasonable, standardized residuals from the weighted least squares regression for each model were calculated. Skewness and kurtosis statistics for the standardized residuals were compared to their 0.05 critical values under the null hypothesis of normality.[†] The skewness and kurtosis statistics are given by:

$$\text{Skewness} = m_3 / m_2^{3/2}$$

and

$$\text{Kurtosis} = (m_4 / m_2^2) - 3$$

where $m_k = (1/N) \sum (\hat{u}_{ij} - \bar{u})^k$, \hat{u}_{ij} is the standardized residual for cell ij , and \bar{u} is the mean of the \hat{u}_{ij} s.

Three procedures were used to check for nonrandom errors. First, in view of the evidence that the log-linear functional form may systematically overstate

^{*} Since the log-linear prediction formula is used for the Box-Cox method when $c = 0$, the log-linear model cannot really be viewed as a special case of the Box-Cox model in terms of prediction.

[†] These tests must be considered approximate, since the standardized residuals will not be independent. If the standardized residuals were to appear highly nonnormal, a possible alternative to the normality assumption and maximization of L might simply be to search for the value of c that minimizes ξ^2 .

actual cell means for cells with the highest average claims, the predicted values and prediction errors for the three cells with the highest predicted values are compared for each model and data set. Second, based on these results and inspection of the prediction errors by cell, additional models are estimated that include interaction terms to mitigate systematic patterns in the errors. Third, the maximum absolute standardized residual for each model is compared to the 0.05 estimated upper bound tabulated by LUND (1975) for this statistic under the normality assumption. This test may provide evidence of whether the error for at least one of the cells cannot be attributed to chance.

With respect to the balance criterion suggested by BAILEY and SIMON (1960), only the linear model will generally produce predicted values that are balanced overall and by class. An estimation method to minimize weighted squared prediction errors subject to balancing constraints later was suggested by BAILEY (1963) and analyzed by JUNG (1968) and AJNE (1974). The motivation for the balance criterion evidently is that it is reasonable to assume full credibility for certain partitions of the data. Within the context of statistical modeling of auto insurance claims, however, there would appear to be no persuasive reason to assume that the deviation of actual experience from the average true pure premium for any particular subset of data or for all exposures is zero (cf WEISBERG and TOMBERLIN (1982)). No attempt is made to compare the models on the basis of the balance criterion.

4. ILLUSTRATIVE RESULTS

The Box and Cox procedure is illustrated in this section using data for two coverages from two countries. The first data set includes numbers of exposures and average claims per exposure for combined compulsory coverages (third-party liability and personal injury protection) in the state of Massachusetts during 1976. These data, which are reported in CHANG and FAIRLEY (1979), are cross-classified for seven driver classes and fifteen territories, yielding a total of 105 cells.* The total number of exposures exceeds 2.4 million with 481 exposures in the smallest cell.

The second data set, which is reported in SAMSON and THOMAS (1984), contains analogous information for a large United Kingdom motor insurance account. The data are for collision coverage during 1977. They are cross-classified by area (rural, semi-urban, and urban), vehicle type (small, medium, and large or powerful), age of driver (17-24, 25-30, and over age 30), and amount of no claim bonus (none, medium, and full), yielding a total of 81 cells. The total number of exposures exceeds 1.7 million with 530 exposures in the smallest cell.

Estimation results for the Massachusetts data under the assumption that d equals one are shown in Table 1. The left-hand side shows results when the

* The driver classes are (1) driver over age 65, (2) driver between age 25 and 65, (3) business use of vehicle, (4) unmarried female driver under age 25, (5) married under age 25, (6) unmarried male non-owner under age 25, and (7) unmarried male owner under age 25

TABLE 1
MODEL COMPARISON FOR MASSACHUSETTS DATA

Statistic	Without Interaction			With Interaction		
	$c = 0$	$c = 1$	$\hat{c}_{\max} = 0.45$	$c = 0$	$c = 1$	$\hat{c}_{\max} = 0.45$
Log-Likelihood	-356.46 ^a	-361.25 ^a	-344.29	-349.49 ^a	-353.84 ^a	-338.85
Mean Squared Error	40.17	18.73	17.73	22.05	16.26	14.46
Mean Absolute Error	2.52	1.99	1.91	2.22	1.94	1.82
Maximum $ \hat{u} $	2.75	3.69 ^b	2.95	2.99	3.95 ^b	3.26
Skewness of \hat{u}	0.30	0.84 ^c	0.52 ^c	0.38	0.73 ^c	0.59 ^c
Kurtosis of \hat{u}	0.04	2.90 ^c	0.27	0.18	2.90 ^c	0.57
Largest \hat{p} s ($\hat{p} - \bar{p}$ is in parentheses)	247.44 (78.47)	168.27 (-0.70)	193.56 (24.59)	184.28 (15.31)	169.89 (0.92)	175.23 (6.26)
	212.49 (69.16)	155.53 (12.02)	172.13 (28.80)	181.12 (30.87)	156.85 (13.52)	159.51 (9.26)
	205.35 (42.27)	154.28 (-8.80)	168.42 (5.34)	168.04 (21.58)	155.71 (-7.37)	155.11 (11.78)

^a Null hypothesis that $c = 0$ or $c = 1$ rejected at 0.05 level.

^b Maximum absolute standardized residual exceeds 0.05 upper bound

^c Null hypothesis of normality rejected at 0.05 level

models are estimated without any interaction terms for territory and driver class.* The results clearly indicate the poor fit in terms of mean squared error and systematic bias for the log-linear model ($c = 0$) that were noted by CHANG and FAIRLEY (1979). While the skewness and kurtosis statistics for the standardized residuals do not reject the normality hypothesis, the three largest \hat{p} s substantially exceed the corresponding cell means.

The linear model ($c = 1$) without interaction terms eliminates the systematic bias for the high-risk cells and provides a substantially better fit than the log-linear model, as was found by Chang and Fairley. However, the standardized residuals for the linear model are poorly behaved. The skewness and kurtosis statistics indicate significant positive skewness and very heavy tails relative to the normal distribution. In addition, the maximum absolute standardized residual exceeds the 0.05 upper bound, and the value of the log-likelihood is less than that for the log-linear model.

The estimated optimal value of c (\hat{c}_{\max}) for the model without interaction terms is 0.45. Application of the likelihood ratio test would result in rejection of the null hypothesis that $c = 0$ or that $c = 1$, but the significant skewness of the standardized residuals for this case suggests caution in interpreting the results of these tests. Both the mean squared error and mean absolute error are lower for $\hat{c}_{\max} = 0.45$ than for $c = 0$ or $c = 1$. Thus, the Box and Cox procedure gives the best overall fit for these data. However, the positive values of $\hat{p} - \bar{p}$ for the largest \hat{p} s also occur for the Box and Cox model, although the problem is much less severe than for the log-linear model.

* The \bar{R}^2 s using the exposure-weighted and c transformed data were about 0.95 or greater for all of the models shown in Tables 1 and 2.

TABLE 2
MODEL COMPARISON FOR UNITED KINGDOM DATA

Statistic	Without Interaction			With Interaction		
	$c = 0$	$c = 1$	$\hat{c}_{\max} = -0.1$	$c = 0$	$c = 1$	$\hat{c}_{\max} = 0.15$
Log-Likelihood	-200.72	-246.22 ^a	-199.56	-168.44 ^a	-218.96 ^a	-165.91
Mean Squared Error	2.75	7.27	2.58	1.32	3.71	1.26
Mean Absolute Error	0.87	1.80	0.84	0.51	0.89	0.48
Maximum $ \hat{u} $	3.36 ^b	2.40	3.28	2.86	2.38	2.87
Skewness of \hat{u}	-0.65 ^c	0.42	-0.58 ^c	-0.24	0.26	-0.30
Kurtosis of \hat{u}	0.64	-0.34	0.52	0.02	-0.27	-0.12
Largest \hat{p} s ($\hat{p} - \bar{p}$ is in parentheses)	87.90 (-0.93)	62.24 (-26.60)	90.74 (1.91)	93.51 (4.68)	66.21 (-22.62)	84.62 (-4.21)
	79.19 (6.42)	58.35 (-14.42)	82.37 (9.60)	82.82 (10.05)	61.12 (-11.65)	76.13 (3.36)
	72.32 (-0.48)	58.09 (-14.71)	73.76 (0.96)	75.03 (2.23)	60.60 (-12.20)	70.64 (-2.16)

^a Null hypothesis that $c = 0$ or $c = 1$ rejected at 0.05 level.

^b Maximum absolute standardized residual exceeds 0.05 upper bound

^c Null hypothesis of normality rejected at 0.05 level

Table 1 also contains the results of estimating the three models when two interaction terms were included in a rough attempt to mitigate the problem of positive values for $\hat{p} - \bar{p}$ for the high-risk classes under the log-linear and Box and Cox models. The first interaction term was for the married under age 25 driver class and the three territories (numbers 13, 14, and 15 in the CHANG and FAIRLEY (1979) paper) with the highest average claims per exposure. The second was for the two male under age 25 driver classes and the same three territories. The inclusion of the interaction terms results in lower mean squared and mean absolute errors for all three models. The reduction is especially pronounced for the log-linear model. The value of \hat{c}_{\max} remains 0.45. The largest \hat{p} s for the Box and Cox model compare favorably with those for the linear model, although \hat{p} still exceeds \bar{p} in each case.

Analogous results for the United Kingdom data, again assuming that d equals one, are shown in Table 2. For the models without interaction terms, the log-linear model substantially outperforms the linear model in terms of predictive accuracy. The value of \hat{c}_{\max} is -0.1 , so that the Box and Cox and log-linear models give similar results. The skewness of the standardized residuals for the log-linear and Box and Cox models is negative and results in rejection of the normality hypothesis at the 0.05 level. In contrast to the results for the Massachusetts data, the linear model appears to produce systematic bias. Each of the three largest \hat{p} s is substantially less than the cell mean. The corresponding results for the log-linear and Box and Cox models provide little evidence of bias.

As noted by SAMSON and THOMAS (1984), there were clear patterns in the prediction errors for the linear and log-linear models without interaction terms. These patterns also were present for the Box and Cox model. Based on their

discussion and on inspection of these patterns, three interaction terms were included. This modification substantially reduced the evidence of nonrandom errors across cells. The first two terms were for drivers in semi-urban and urban areas with powerful vehicles, respectively. The third was for urban drivers of powerful vehicles with full no claim bonus.

The results shown in Table 2 indicate a substantially better fit when these interaction terms are included, and the normality hypothesis for the standardized residuals cannot be rejected for any of the models. The Box and Cox procedure, with \hat{c}_{\max} equal to 0.15, produces the lowest mean squared error and mean absolute error. The log-linear model is a close competitor, however. Application of the likelihood ratio test indicates rejection of the hypothesis that $c = 0$ or that $c = 1$. The three largest \hat{p} s for the Box and Cox model are approximately equal to the cell means and are lower in magnitude than are those for the log-linear model. The linear model still produced \hat{p} s that are considerably less than the cell means.

The Box and Cox procedure also was used to estimate the models with the interaction terms for both data sets without constraining d to equal one. Selected estimation results are set forth below.

Massachusetts data: $L = -165.017$, $\hat{c} = 0.15$, $\hat{d} = 1.2$

United Kingdom data: $L = -338.224$, $\hat{c} = 0.5$, $\hat{d} = 1.15$.

The values of L and \hat{c} are quite similar to those shown in Tables 1 and 2. In neither case is the null hypothesis that $d = 1$ rejected at the 0.05 level on the basis of a likelihood ratio test. Moreover, the predicted values and standardized residuals for the unconstrained cases were very similar to those for the constrained cases.

5. CONCLUSION

The results suggest that estimation and testing for functional form in automobile insurance pure premium models using the Box and Cox procedure may provide more accurate predictions than simply assuming either a linear or log-linear model. Future work should apply this technique to additional data to determine whether it is able consistently to improve predictive accuracy. If possible, this work should evaluate the accuracy of model predictions using data for a period subsequent to the estimation period. It would be desirable to extend the method to models of claim frequency, severity, or both. Consideration also might be given to using numerical methods to derive consistent predictions for the Box and Cox model. Moreover, if the procedure is applied to estimate pure premium models and the normality assumption is clearly violated, it might be desirable to analyze performance of the method when the transformation parameter is chosen on the basis of a minimum squared error criterion.

The results also highlight the need to refine procedures for selecting interaction terms for pure premium models. Judicious selection of even a few interaction terms may be able significantly to improve overall predictive accuracy and mitigate or possibly eliminate bias in predicted values. Analyses of issues such as why

the ratio of average claims for high-risk drivers to average claims for low-risk drivers in urban areas tends to be less than the corresponding ratio in rural areas may provide guidance in this regard. Alternatively, additional analysis of the ability of stepwise or other regression selection methods to produce models with improved predictive accuracy is needed, and the feasibility and potential advantages of using the Box and Cox procedure in conjunction with such methods should be explored.

ACKNOWLEDGEMENT

The author wishes to thank A. Gisler and an anonymous referee for helpful comments. Thanks also are due to Jaewan Her and Boem-ha Jee for competent research assistance and to Danny Samson and Howard Thomas for providing their data.

REFERENCES

- AJNE, B (1974) A Note on the Multiplicative Ratemaking Model *ASTIN Bulletin* 8, 144-153
- ALBRECHT, P (1983) Parametric Multiple Regression Risk Models Theory and Statistical Analysis *Insurance, Mathematics and Economics* 2, 49-66
- ALMER, B. (1957) Risk Analysis in Theory and Practical Statistics *15th International Congress of Actuaries* 2, 314-353
- AMEMIYA, T (1973) Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of its Expectation *Journal of the American Statistical Association* 68, 928-934
- BAILEY, R (1963) Insurance Rates with Minimum Bias *Proceedings of the Casualty Actuarial Society* 50, 4-11
- BAILEY, R and SIMON, L (1960) Two Studies in Automobile Insurance Ratemaking. *Proceedings of the Casualty Actuarial Society* 47, 1-19
- BOX, G and COX, D (1964) An Analysis of Transformations *Journal of the Royal Statistical Society B* 26, 211-243
- CHAMBERLAIN, C (1980) Relativity Pricing through Analysis of Variance *Pricing Property and Casualty Insurance Contracts*, 1980 CAS Discussion Paper Program, San Juan, Puerto Rico
- CHANG, L and FAIRLEY, W (1978) An Estimation Model for Multivariate Insurance Rate Classification *Automobile Insurance Risk Classification. Equity & Accuracy* Boston, Mass · Massachusetts Division of Insurance
- CHANG, L. and FAIRLEY, W (1979) Pricing Automobile Insurance Under Multivariate Classification of Risks Additive versus Multiplicative *Journal of Risk and Insurance* 46, 75-98.
- COUTTS, S. (1984) Motor Insurance Rating, An Actuarial Approach *Journal of the Institute of Actuaries* 111, 87-148.
- EEGHEN, J, GREUP, E and NIJSSSEN, J (1983) *Ratemaking* Surveys of Actuarial Studies No. 2 Nationale-Nederlanden N V Research Department, Rotterdam, the Netherlands.
- FAIRLEY, W, TOMBERLIN, T and WEISBERG, H. (1981) Pricing Automobile Insurance Under a Cross-Classification of Risks Evidence from New Jersey *Journal of Risk and Insurance* 48, 505-514.
- FREIFELDER, L (1984) Estimation of Classification Factor Relativities. A Modeling Approach Paper presented at the 1984 Risk Theory Seminar, Los Angeles, California
- HOLMES, R (1970) Discriminatory Bias in Rates Charged by the Canadian Auto Insurance Industry *Journal of the American Statistical Association* 65, 108-122
- JOHNSON, P and HEY, G (1971) Statistical Studies in Motor Insurance *Journal of the Institute of Actuaries* 97, 201-232
- JUNG, J (1968) On Automobile Insurance Ratemaking *ASTIN Bulletin* 5, 41-48
- LAHIRI, K and EGY, D (1981) Joint Estimation and Testing for Functional Form and Heteroskedasticity. *Journal of Econometrics* 15, 299-307.
- LUND, R. (1975) Tables for an Approximate Test for Outliers in Linear Models *Technometrics* 17, 473-476

- NELSON, H and GRANGER, C (1979) Experience with Using the Box-Cox Transformation when Forecasting Economic Time Series *Journal of Econometrics* **10**, 57-69.
- OBERHOFER, W. and KMENTA, J (1974) A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models *Econometrica* **42**, 579-590
- SAMSON, D and THOMAS, H (1984) Claims Modeling in Auto Insurance Department of Business Administration, University of Illinois, Urbana, Illinois
- SANT, D (1980) Estimating Expected Losses in Auto Insurance *Journal of Risk and Insurance* **47**, 133-151
- SEAL, H (1968) The Use of Multiple Regression in Risk Classification Based on Proportionate Losses *18th International Congress of Actuaries* **2**, 659-664
- WEISBERG, H. and TOMBERLIN, T (1982) A Statistical Perspective on Actuarial Methods for Estimating Pure Premiums from Cross-Classified Data *Journal of Risk and Insurance* **49**, 539-563
- WEISBERG, H, TOMBERLIN, T and CHATTERJEE, S (1984) Predicting Insurance Losses Under Cross-Classification A Comparison of Alternative Approaches *Journal of Business and Economic Statistics* **2**, 170-178
- WILCKEN, C (1971) Comment on 'Discriminatory Bias in Rates Charged by the Canadian Automobile Insurance Industry' *Journal of the American Statistical Association* **66**, 289-291

SCOTT E. HARRINGTON

Insurance Department, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA

