THE PERFORMANCE OF ALTERNATIVE MODELS FOR FORECASTING AUTOMOBILE INSURANCE PAID CLAIM COSTS

J. David Cummins and Alwyn Powell

1. INTRODUCTION

When setting or modifying insurance rates on the basis of observed data, adjustments should be made for inflation between the average claim date of the observation period and the average claim date covered by policies to which the new rates will apply. In automobile insurance ratemaking in the United States, the adjustment is made as follows [1]:

$$(1) \qquad L_a = L_0 \times (1 + \widehat{\beta})^k$$

where L_a = trended losses and loss adjustment expenses incurred,
$\quad\ L_0$ = observed losses and loss adjustment expenses incurred during the experience period,
$\quad\ \widehat{\beta}$ = the predicted rate of inflation per quarter, and
$\quad\ k$ = the number of quarters for which losses are being trended.

The inflation coefficient, β, is an exponential growth factor that is estimated using the following equation [2]:

$$(2) \qquad \ln \bar{y}_t = \alpha + \beta t + \varepsilon_t$$

where $\bar{y}_t$ = average paid claim costs for the year ending in quarter t,
$\quad\ t$ = time, in quarters, $t = 1, 2, \ldots, 12$, and
$\quad\ \varepsilon_t$ = a random disturbance term.

The estimation period for the equation is the twelve quarters immediately preceding the first forecast period, and the estimation is conducted utilizing ordinary least squares. Annual averages of claim costs are used as the dependent variable to smooth random and seasonal fluctuations.

The use of equation (2) can be expected to lead to two major problems:

1. Since the equation merely extrapolates past inflation rates into the future, the forecasts are likely to be seriously inaccurate when the inflation rate in the forecast period is substantially different from that in the preceding twelve quarters.

---

[1] In some cases, an adjustment also is made for frequency. The formula in such cases is: $L_a = L_0(1 + \widehat{\beta}_s)^k \cdot (1 + \widehat{\beta}_f)^k$, where $\widehat{\beta}_s$ = the predicted rate of change in severity and $\widehat{\beta}_f$ = the predicted rate of change in frequency. This article is limited to an analysis of severity, which is the predominant source of claims cost inflation.

[2] Trend factors are discussed in Cook (1970). When frequency trend factors are employed, they also are obtained through exponential trending.

2. Forecasts based on four quarter moving averages are likely to respond slowly to changing inflation rates. Thus, premium rates will be unresponsive, contributing to the underwriting cycle.

An alternative approach to forecasting average paid claim costs is to use an econometric model which predicts claim costs as a function of relevant economic variables. This approach has become feasible in recent years as progress has been made in developing forecasting models for major economic aggregates in most industrialized nations [3]. The purpose of this article is to develop and test the accuracy of econometric models for predicting automobile liability insurance paid claim costs. The estimation and testing are conducted using claim cost data for the United States.

2. EQUATION SPECIFICATION AND METHODOLOGY

Among the advantages of the exponential trend model (equation (2)) are its familiarity and simplicity. These are important qualities because rate increases in most states must receive regulatory approval. Since regulators are likely to be suspicious of highly technical models, any methodology proposed as a replacement for exponential trending should be as simple as possible. Hence, the approach taken in this article is to begin by utilizing ordinary least squares (OLS) to estimate equations with the simplest possible specification. Tests then are conducted to determine whether the resulting equations depart from the assumptions underlying the OLS technique. Where departures are found, the equations are reestimated with the appropriate adjustments.

The three basic specifications used in the research are the following:

$$(3) \qquad y_t = \alpha + \beta x_t + \varepsilon_t$$

$$(4) \qquad \ln y_t = \alpha + \beta \ln x_t + \varepsilon_t$$

$$(5) \qquad y_t = \alpha + \beta x_t + \beta_2 y_{t-1} + \varepsilon_t$$

where y_t = average paid claim costs in quarter t,
x_t = a price or wage index, and
ε_t = the random disturbance term.

Linear and log-linear equations in which two economic indices appear as explanatory variables also were estimated, and dummy variables for seasonality were tested with each type of equation. Since the same estimation methods apply to (3) and (4), the examples pertaining to these equations in the following discussion are based only on equation (3).

[3] Models of the U.S. economy are reviewed in KLEIN and BURMEISTER (1976). For models applicable to other countries, see WAELBROECK (1976).

The classical ordinary least squares model incorporates the following assumptions about ε_t [4]:

$$(6) \qquad\qquad E(\varepsilon_t) = 0$$

$$(7) \qquad\qquad E[(x_t - \bar{x}_t)\varepsilon_t] = 0$$

$$(8) \qquad\qquad E(\varepsilon_t^2) = \sigma^2$$

$$(9) \qquad E(\varepsilon_t \cdot \varepsilon_{t-i}) = 0, \text{ for } i = 1, 2, \ldots t.$$

If the assumptions are satisfied, ordinary least squares yields parameter estimates for equations (3) and (4) that are best linear unbiased, i.e., that have the smallest variances among all linear unbiased estimators. It is not necessary to assume normality in order to achieve this result, although the usual tests of significance for the coefficients will be correct only asymptotically if normality is not present. Even if all of the assumptions are satisfied, the ordinary least squares estimators for equation (5) are biased although they are consistent and asymptotically efficient.

As assumptions (6) and (7) rarely cause problems, attention is focused on assumptions (8) and (9), i.e., homoskedasticity and the absence of serial correlation, respectively. If either assumption is violated, ordinary least squares estimators of the parameters in equations (3) and (4) are unbiased and consistent but are inefficient both in finite samples and asymptotically. For equation (5), violation of assumption (8) implies that ordinary least squares estimators are consistent but asymptotically inefficient, while violation of assumption (9) renders the OLS estimators both inconsistent and asymptotically inefficient.

Because of the potential importance of assumptions (8) and (9) for forecast accuracy, tests are conducted to determine whether they are satisfied in the estimated relationships. For equations (3) and (4), the Durbin-Watson d statistic is used to test for serial correlation, while the hypothesis of homoskedasticity is tested using a procedure suggested by GOLDFELD and QUANDT (1965) [5]. Because the Durbin-Watson d statistic is inappropriate for equations with lagged dependent variables among the explanatory variables, an alternative procedure suggested by DURBIN (1970) is used to test for serial correlation in equation (5) [6]. The Goldfeld-Quandt heteroskedasticity test also is applied to equation (5).

---

[4] The discussion of estimation theory in this section draws heavily on KMENTA (1971).
[5] The Goldfeld-Quandt test has been shown to be at least as powerful as other major heteroskedasticity tests. See HARVEY and PHILLIPS (1974) and HARRISON and McCABE (1979).
[6] DURBIN proposed two tests—one based on the statistic h and the other on regressions involving the OLS residuals of the equation being tested. The latter test is used here because the h statistic has been shown to be biased in small samples. See SPENCER (1975).

If the tests indicate that autocorrelation is present in equation (3) or (4), the approach is to hypothesize that $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, where u_t satisfies (6) through (9). The equation then is reestimated by solving the following problem:

$$(10) \qquad \underset{(\alpha,\beta,\rho)}{\text{Minimize}} \ \sum_{t=2}^{T} \ [y_t - \rho y_{t-1} - \alpha(1 - \rho) - \beta(x_t - \rho x_{t-1})]^2$$

Minimizing (10) is asymptotically equivalent to maximum likelihood estimation if u_t is normally distributed [7].

A method suggested by GLEJSER (1969) is used to adjust for heteroskedasticity. This technique is based on the hypothesis that $\varepsilon_t = P(z)w_t$, where $P(z)$ is a function of a mathematical variable z, and w_t satisfies (6) through (9). The variable z is taken to be (one of) the independent variable(s) in the regression, in this case x_t. The estimation is conducted as follows:

(1) Estimate equation (3) or (4) using OLS and obtain the residuals, $\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$.

(2) Regress $|\hat{\varepsilon}_t|$ on simple functions of the independent variable such as $\gamma_0 + \gamma_1 x_t$, $\gamma_0 + \gamma_1 x_t^2$, $\gamma_0 + \gamma_1 x_t^{1/2}$, etc. and choose as the functional form $\hat{P}(x_t)$ the equation with the highest explanatory power (R^2).

(3) Transform the regression equation by dividing each term by $\hat{P}(x_t)$. E.g., if the best regression for $|\varepsilon_t|$ is $\hat{\gamma}_0 + \hat{\gamma}_1 x^2$, the transformed variables are equal to $y_t' = y_t/(\hat{\gamma}_0 + \hat{\gamma}_1 x_t^2)$, $x_t' = x_t/(\hat{\gamma}_0 + \hat{\gamma}_1 x_t^2)$, and $c_t' = 1.0/(\hat{\gamma}_0 + \hat{\gamma}_1 x_t^2)$, where c_t' is the transformed constant term.

(4) Estimate specification (3) or (4) using the transformed variables and OLS. For example, equation (3) becomes: $y_t' = \alpha c_t' + \beta x_t' + w_t$.

If the hypothesis is correct, the error term of the transformed equation satisfies (6) through (9). However, the correction will be only approximate for finite sample sizes.

When the error terms in equation (3) or (4) are characterized by both autocorrelation and heteroskedasticity, the estimation problem is more difficult. The ideal solution would be to use a generalized nonlinear estimation procedure which would correct simultaneously for both problems. Since computer programs to perform this type of estimation were not readily available to the authors, two alternative approaches were adopted, based on extensions of Glejser's method.

Method 1. The following error structure is hypothesized: $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, where $u_t = P(x_t)w_t$; $E(u_t\varepsilon_{t-i}) = E(u_t u_{t-i}) = E(w_t w_{t-i}) = 0$, for $i = 1, 2, \ldots, t$; and $E(w_t^2) = \sigma^2$. Under this hypothesis, the first step in the estimation is to solve the minimization problem in expression (10). The absolute values of the estimated residuals from this equation then are regressed on various

---

[7] Minimization was conducted using the Time Series Processor (TSP) program. See HALL, HALL, and BECKETTI (1977).

functions of x_t to determine $\hat{P}(x_t)$. Finally, OLS is used to estimate the following equation:

$$(11) \qquad \frac{y_t - \hat{\rho}y_{t-1}}{\hat{P}(x_t)} = \frac{\alpha(1 - \hat{\rho})}{\hat{P}(x_t)} + \frac{\beta(x_t - \hat{\rho}x_{t-1})}{\hat{P}(x_t)} + \frac{u_t}{\hat{P}(x_t)}$$

where $\hat{\rho}$ = the estimate of ρ obtained in the first stage by solving the minimization problem in (10), and

$\hat{P}(x_t)$ = the functional form of x_t which provides the highest explanatory power in regressions of the absolute residuals from (10).

Method 2. In this case, the error structure is assumed to be: $\varepsilon_t = P(x_t)[\rho u_{t-1} + w_t]$, where $E(u_{t-i}w_t) = E(w_t w_{t-i}) = 0$, $i = 1, 2, \ldots, t$; and $E(w_t^2) = \sigma^2$. The estimation procedure is to obtain OLS estimates of α and β, define $\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$, and regress $|\varepsilon_t|$ on various functions of x_t. Denoting the best of these functions as $\hat{P}(x_t)$, the final step is to minimize the following expression with respect to α, β, and ρ:

$$(12) \qquad \sum_{t=2}^{T} \left[\frac{y_t}{\hat{P}(x_t)} - \frac{\rho y_{t-1}}{\hat{P}(x_t)} - \frac{\alpha(1 - \rho)}{\hat{P}(x_t)} - \frac{\beta(x_t - \rho x_{t-1})}{\hat{P}(x_t)} \right]^2 .$$

Equations estimated through both methods were used in the forecasting experiments with the choice between the two based on forecast accuracy.

Efficient estimation of equations with lagged dependent variables among the regressors when the residuals are autocorrelated or heteroskedastic is a difficult econometric problem requiring specialized computer programs not available to the authors. Accordingly, no adjustments were made when equation (5) was characterized by these problems, and the OLS versions were used in the forecasting experiments.

3. THE DATA AND THE FORECASTING EXPERIMENTS

The data used in estimating the forecasting equations are automobile bodily injury and property damage liability insurance average paid claim costs reported by the Insurance Services Office (ISO). The ISO is the leading rating bureau in the United States, maintaining automobile insurance data for more than one hundred contributing insurance companies. Pooled data from all contributing companies are used by ISO to develop trend factors that affect the rates of a subset of contributors (others price independently). A credibility weighted average of state and national data is used in the trending process for rates applicable to a particular state. The forecasts in this article are based on the national averages.

The data period for automobile property damage liability claim costs is the first quarter of 1954 (1954.1) to the second quarter of 1978 (1978.2). For

bodily injury liability, the period is 1964.1 to 1978.2. The property damage figures are for total limits coverage, while the bodily injury data are for $ 10,000 basic limits coverage (i.e., claims in excess of $ 10,000 appear in the averages as $ 10,000). Each series is obtained by dividing the total dollar amount of claim payments in a given quarter by the total number of claims in the quarter. The claims cost series, converted to index form, are presented in the Appendix [8].

Several price and wage indices were tested as independent variables. Among those with the highest explanatory power are the implicit price deflator for gross national product, the implicit price deflator for autos and parts, and wage indices for the service sector and for the private sector of the economy. Actual values of the deflators and indices are available in published U.S. government sources [9], while predicted values were obtained from Wharton Econometric Forecasting Associates (WEFA), one of the leading U.S. forecasting firms [10].

The initial screening of predictor variables was based on summary regression statistics such as the coefficient of determination (R^2) [11]. After selecting the two or three best variables, subsequent evaluations of variables and specifications were based on predictive accuracy. The approach taken in the experiments was to forecast average paid claim costs for periods of eight quarters beginning at one-year intervals from 1971 through 1977 [12]. The third quarter of each year was arbitrarily selected as the initial forecast period. The goal in the experiments was to reproduce as accurately as possible the results that would have been obtained if the forecasting equations actually had been employed in the periods under consideration. Accordingly, the econometric models were estimated using actual values of the dependent and independent variables for the period ending in the quarter prior to the initial forecast period, and the forecasts were generated using predicted values of the independent variables.

---

[8] Transforming data into indices affects the coefficients of any regressions based on the transformed data but not the values of the essential test statistics. KMENTA (1971), pp. 377-379.

[9] The full title of the autos and parts deflator is the implicit deflator, personal consumption expenditures, durables, autos and parts. It and the GNP deflator are available in the *Survey of Current Business* (Washington, D.C.: U.S. Government Printing Office, monthly). The wage indices are compensation per man hour, private sector and compensation per man hour, service sector. They are constructed from series available in the *Survey of Current Business* and in *Employment and Earnings* (Washington, D.C.: U.S. Government Printing Office, monthly).

[10] WEFA forecasts are released quarterly. A summary version of the forecast is presented each quarter in *The Wharton Magazine*, which is published at the University of Pennsylvania, Philadelphia, Pa.

[11] R^2 is not a good indicator of predictive accuracy. However, it was considered acceptable as an initial screening statistic.

[12] Data availability limited the last forecast period to four quarters.

The accuracy of the forecasts is determined by reference to the following statistics:

$$(13) \qquad \text{TPCE} = \left(\frac{\hat{y}_8 - y_8}{y_0}\right) \times 100$$

$$(14) \qquad \text{MAPE} = \sum_{t=1}^{8} \frac{|y_t - \hat{y}_t|}{y_t} \times 100$$

$$(15) \qquad \text{RMSPE} = 100 \times \sqrt{\sum_{t=1}^{8} \left(\frac{y_t - \hat{y}_t}{y_t}\right)^2}$$

where TPCE = total predicted change error,
       MAPE = mean absolute percentage error,
       RMSPE = root mean square percentage error,
       $\hat{y}_t$ = the predicted value of average paid claim costs in quarter t of the forecast period,
       y_t = actual average paid claim costs in quarter t of the forecast period, and
       y_0 = the actual value of average paid claim costs in the quarter immediately preceding the forecast period.

The TPCE is the most important measure of forecasting accuracy for ratemaking purposes because it measures the percentage by which predicted claim costs exceed (fall short of) actual claim costs for the period to which the new rates would apply.

Tests with the ordinary least squares versions of equations (3), (4), and (5) revealed little difference among alternative predictor variables in terms of goodness-of-fit or forecast accuracy. Since the private sector wage rate performed at least as well as the other variables during most forecast periods, it is used exclusively in the reported regressions. Equations with more than one economic explanatory variable were not much better than the univariate equations in terms of goodness-of-fit and, in fact, often yielded unrealistic signs for the coefficients because of multicollinearity. Accordingly, the results presented below are based on equations with one economic regressor. The best forecasts were obtained with equations estimated over the maximum available observation period.

4. RESULTS: AUTOMOBILE PROPERTY DAMAGE LIABILITY PAID CLAIM COSTS

For automobile property damage liability paid claim costs, the OLS version of equation (5) yielded forecasts which generally were more accurate than those from the OLS versions of equations (3) and (4). Tests revealed that all three specifications are subject to autocorrelation and heteroskedasticity in nearly

every estimation period [13]. Since it was not feasible to adjust equation (5) for these problems due to program limitations, the next-best OLS model, equation (3), was chosen for adjustment. The exponential trend equations and the OLS version of equation (5) are presented in Table 1. Equation (3), adjusted for autocorrelation, is shown in the top panel of Table 2, while the bottom panel shows this specification adjusted for both autocorrelation and heteroskedasticity according to Method 1 [14]. Dummy variables for seasonality are included in the equations presented in Table 2 but not in the OLS version of equation (5) because preliminary results indicated that the inclusion of seasonal dummies in the OLS equations leads to less accurate forecasts.

The forecast results for property damage liability paid claim costs are shown in Table 3. The table gives the TPCE, the MAPE, and the RMSPE for the exponential trend model (ET), the OLS model with lagged endogenous regressor (LE), equation (3) adjusted for autocorrelation (AR), and the same specification adjusted for both autocorrelation and heteroskedasticity (ARH). To provide a benchmark for analyzing the forecast errors, a shift index to measure unanticipated inflation has been included in the table. The index is the ratio of the (quarterly) geometric mean rate of claims cost inflation during each forecast period to the geometric mean inflation rate during the preceding twelve quarters.

In terms of TPCE, all three econometric models outperform the exponential trend model in five of the seven forecast periods. As expected, the performance of the trend model is worst when unanticipated inflation is highest. Among the econometric models, those adjusted for both autocorrelation and heteroskedasticity do best during periods characterized by high unanticipated inflation (e.g., the forecast periods beginning in 1973.4, 1974.4, and 1975.4). The OLS model also does quite well during these periods, suggesting its use as a relatively simple replacement for the exponential trend model.

5. RESULTS: AUTOMOBILE BODILY INJURY LIABILITY PAID CLAIM COSTS

The exponential time trend and best OLS equations for automobile bodily injury liability claim costs are presented in Table 4. In this case, equation (3) is the best OLS model. The hypothesis of homoskedasticity is rejected for this model during the last five estimation periods. The Durbin-Watson test indicates the presence of autocorrelation during the last four periods and gives inconclusive results for the period ending in 1973.2. Since the behavior of the test statistics when autocorrelation and heteroskedasticity are present simultaneously is not clear, the equations for all periods were reestimated

---

[13] All significance test results reported in this article are at the 5 percent level.

[14] For the first two sample periods, $\hat{P}(x_t) = \hat{\gamma}_0 + \hat{\gamma}_1/x_t$; while for the last five periods $\hat{P}(x_t) = \hat{\gamma}_0 + \hat{\gamma}_1 x_t^2$.

TABLE 1: EXPONENTIAL TREND AND LAGGED ENDOGENOUS PREDICTION EQUATIONS FOR AUTOMOBILE PROPERTY DAMAGE LIABILITY AVERAGE PAID CLAIM COSTS

Equation for Forecast Beginning in	$ln\,\hat{y}_t = \hat{\alpha} + \hat{\beta}t$			$y_t = \hat{\alpha} + \hat{\beta}_1 w_t + \hat{\beta}_2 y_{t-1}$					
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2	D [a]	GQ [b]
1971.3	5.3935 (1822.44)	0.0236 (58.740)	.997	−11.8994 (−2.506)	11.5062 (2.136)	.885015 (12.793)	.995	−6.75	2.54 (29,29)
1972.3	5.5068 (1065.12)	0.0194 (27.675)	.987	−11.1340 (−2.198)	12.5708 (2.220)	.860157 (11.992)	.995	−3.96	4.52 (31,31)
1973.3	5.6113 (846.148)	0.0132 (14.656)	.956	−10.9235 (−2.349)	12.5643 (2.326)	.858908 (12.335)	.996	−4.21	4.98 (33,33)
1974.3	5.6756 (1323.17)	0.0116 (19.905)	.975	−5.8705 (−1.366)	8.0022 (1.529)	.908715 (13.119)	.996	−4.67	6.86 (35,35)
1975.3	5.6997 (828.777)	0.0163 (17.426)	.968	−9.4807 (−2.898)	12.3019 (2.904)	.854314 (14.685)	.996	−3.31	7.26 (37,37)
1976.3	5.7499 (721.499)	0.0220 (20.314)	.976	−8.5855 (−2.982)	11.5338 (2.924)	.862572 (15.620)	.997	−3.50	6.63 (39,39)
1977.3	5.8099 (1705.41)	0.0254 (54.842)	.997	−9.9913 (−3.533)	13.2348 (3.280)	.840751 (14.754)	.997	−4.04	6.97 (41,41)

Key: y_t = average paid claim costs in quarter t; $\hat{y}_t$ = average paid claim costs for the year ending in quarter t; t = time, in quarters; and w_t = private sector wages per man-hour in quarter t.

Note: The estimation for each time trend equation is the twelve quarters preceding the first forecast period. For each geometric lag equation, the estimation period begins in 1954.1 and ends in the quarter preceding the first forecast period. The numbers in parentheses below the estimated coefficients are t-statistics.

[a] This column presents results of a test suggested by DURBIN (1970) for autocorrelation in models with lagged endogenous explanatory variables. Durbin suggests two tests, one based on a statistic h, and the other based on a regression involving the residuals of the OLS equation being tested. The latter test is used here. Since it is a t-test with more than 30 degrees of freedom, the standard normal table can be used to test significance. At the 5 percent level, the hypothesis of no autocorrelation is rejected for every equation shown in this section of the table.

[b] This column presents the GOLDFELD-QUANDT (1965) test statistics for heteroskedasticity. The hypothesis of homoskedasticity is rejected if the statistic value exceeds F_{α, ν_1, ν_2}. The degrees of freedom parameters, ν_1 and ν_2 appear in parentheses below the statistic. At the 5 percent level, the hypothesis is rejected for every equation shown in this section of the table.

J. DAVID CUMMINS AND ALWYN POWELL

TABLE 2: AUTOREGRESSIVE EQUATIONS FOR AUTOMOBILE PROPERTY DAMAGE LIABILITY
AVERAGE PAID CLAIM COSTS:
EQUATIONS IN LOWER PANEL ADJUSTED FOR HETEROSKEDASTICITY

Equation for Forecast Beginning in	$y_t = \hat{\rho}y_{t-1} + \hat{\alpha}(1-\hat{\rho}) + \hat{\beta}(w_t - \hat{\rho}w_{t-1}) + \hat{\gamma}_1 S_1 + \hat{\gamma}_2 S_2 + \hat{\gamma}_3 S_3$					
	$\hat{\beta}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\rho}$	R^2
1971.3	85.6807 (13.780)	−9.3200 (−11.828)	−2.3364 (−2.960)	−5.9115 (−7.464)	.950495 (64.735)	.9983
1972.3	77.2711 (21.109)	−9.8140 (−10.518)	−2.7595 (−2.940)	−5.8231 (−6.211)	.924875 (50.785)	.9980
1973.3	76.0740 (23.796)	−10.2579 (−10.725)	−2.7396 (−2.835)	−5.5684 (−5.784)	.918491 (49.380)	.9982
1974.3	72.2098 (29.452)	−10.6073 (−9.390)	−3.0991 (−2.688)	−5.6290 (−4.938)	.897217 (38.963)	.9977
1975.3	72.7631 (31.229)	−10.8620 (−10.101)	−3.5125 (−3.196)	−5.7196 (−5.238)	.895484 (39.829)	.9981
1976.3	72.4086 (35.250)	−10.7907 (−10.588)	−3.5110 (−3.359)	−5.6970 (−5.489)	.893974 (41.125)	.9985
1977.3	73.4537 (35.232)	−11.5857 (−11.108)	−3.6506 (−3.396)	−6.1675 (−5.770)	.892990 (40.288)	.9987
1971.3	89.6291 (37.599)	−14.1917 (−17.047)	−0.7009 (−0.879)	−6.4039 (−7.4297)	.950495	.9751
1972.3	77.9910 (33.828)	−10.7988 (−10.874)	−2.6878 (−2.778)	−5.7983 (−5.938)	.924875	.8983
1973.3	75.5438 (38.015)	−9.4062 (−11.164)	−2.8493 (−3.498)	−5.7251 (−6.979)	.918491	.8030
1974.3	71.0324 (43.322)	−8.3723 (−10.736)	−3.1814 (−4.244)	−5.9034 (−7.786)	.897217	.5889
1975.3	72.0044 (45.473)	−9.3375 (−10.980)	−3.3019 (−4.012)	−5.9526 (−7.179)	.895484	.7483
1976.3	72.2156 (49.337)	−10.1239 (−11.175)	−3.3716 (−3.826)	−5.8504 (−6.614)	.893974	.8724
1977.3	72.7016 (51.657)	−10.3194 (−11.635)	−3.4609 (−4.016)	−6.1045 (−7.050)	.892990	.8612

Key: y_t = average paid claim costs in quarter t; w_t = private sector wages per man-hour in quarter t; and S_i = 1.0 in quarter i of each year, o otherwise.
Note: The estimation period for each equation begins in 1954.1 and ends in the quarter preceding the first forecast period. The number in parentheses below each estimated coefficient is the ratio of the coefficient to its estimated standard deviation. The constant term α was estimated in an earlier version of the equations but was found to be insignificant.

TABLE 3: RESULTS OF EIGHT QUARTER FORECASTS OF AUTOMOBILE PROPERTY DAMAGE LIABILITY AVERAGE PAID CLAIM COSTS

First Period of Forecast	Total Predicted Change Errors [a]				Mean Absolute Percent. Errors [a]				Root Mean Square Percent. Errors [a]				Shift Index [b]
	ET	LE	AR	ARH	ET	LE	AR	ARH	ET	LE	AR	ARH	
1971.3	10.8%	5.4%	3.7%	4.3%	6.0%	3.8%	2.5%	2.4%	6.9%	4.4%	2.7%	2.7%	51%
1972.3	5.4	7.3	6.0	6.3	5.7	2.4	1.9	1.9	5.9	3.2	2.4	2.5	61
1973.3	−10.3	1.8	0.3	0	1.1	3.0	2.1	2.0	1.7	3.7	2.9	2.9	126
1974.3	−15.6	−2.4	0.6	0	5.6	2.9	1.2	1.4	6.4	3.1	1.6	1.8	201
1975.3	−8.7	1.6	0.5	0	5.6	1.7	1.8	1.5	5.9	1.9	2.0	1.8	150
1976.3	−6.9	−4.3	−5.0	−5.0	2.9	1.9	2.0	2.1	3.2	2.4	2.5	2.5	138
1977.3 [c]	−1.6	−3.0	−3.4	−3.6	1.2	1.9	2.0	2.2	1.2	2.2	2.3	2.5	115
Absolute Averages	8.5%	3.7%	2.8%	2.7%	4.0%	2.5%	1.9%	1.9%	4.5%	3.0%	2.3%	2.4%	

Key: ET = exponential trend model, LE = econometric model, AR = autoregressive model with lagged endogenous regressor, ARH = autoregressive model without heteroskedasticity adjustment, ARH = autoregressive model with heteroskedasticity adjustment.

Note: The econometric forecasts (columns labelled LE, AR, and ARH) are based on predicted values of the private sector wage rate promulgated by Wharton Econometric Forecasting Associates, Inc. The equations were estimated using revised values of the wage rate series, which differ slightly in most cases from the published values available at the time the Wharton forecasts were made. Thus, the predicted value of the wage rate, $\hat{w}_t$, for use in the claim cost forecasts is $\hat{w}_t = w_{0c} \cdot \hat{w}_{tw}/w_{0p}$, where w_{0c} = the current (revised) value of w for the period preceding the first forecast period, w_{0p} = the (preliminary) value of w for the period preceding the first forecast period at the time the Wharton forecast was made, and $\hat{w}_{tw}$ = the Wharton forecast of w for period t.

[a] Total predicted change error (TPCE), mean absolute percentage error (MAPE), and root mean square percentage error (RMSPE) are defined in equations (13), (14), and (15) in the text.

[b] The shift index is computed as follows: $I = [(y_8/y_0)^{1/8} - 1] \div [(y_0/y_{-12})^{1/12} - 1]$, where y_8 = actual average paid claim costs in the last quarter of the forecast period, y_0 = actual claim costs in the quarter preceding the forecast period, and y_{-12} = actual claim costs thirteen quarters prior to the forecast period.

[c] This forecast period runs from 1977.3 through 1978.2.

TABLE 4: EXPONENTIAL TREND AND LINEAR OLS PREDICTION EQUATIONS FOR AUTOMOBILE BODILY INJURY LIABILITY AVERAGE PAID CLAIM COSTS

Equation for Forecast Beginning in	$\ln \bar{y}_t = \hat{\alpha} + \hat{\beta} t$			$\bar{y}_t = \hat{\alpha} + \hat{\beta} w_t$				
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2	DW [a]	GQ [b]
1971.3	6.9943 (1868.45)	0.0147 (28.938)	.988	−147.603 (−2.929)	340.042 (24.356)	.955	2.088	1.35 (10,10)
1972.3	7.0458 (844.845)	0.0171 (15.070)	.958	−85.923 (−1.884)	321.892 (26.432)	.956	1.827	1.93 (12,12)
1973.3	7.1303 (505.193)	0.0116 (6.053)	.786	−14.908 (−0.360)	301.299 (28.304)	.957	1.461	2.29 (14,14)
1974.3	7.1964 (603.596)	0.0094 (5.772)	.769	27.2017 (0.712)	289.316 (30.657)	.959	1.356*	3.25 (16,16)
1975.3	7.1937 (1210.95)	0.0174 (21.525)	.979	104.425 (2.907)	268.313 (31.695)	.958	1.208*	2.93 (18,18)
1976.3	7.2514 (1125.23)	0.0200 (22.812)	.981	107.879 (3.590)	267.402 (39.618)	.970	1.255*	3.37 (20,20)
1977.3	7.3304 (1162.57)	0.0202 (23.549)	.982	112.189 (4.347)	266.264 (48.278)	.978	1.399*	2.67 (22,22)

Key: y_t = average paid claim costs in quarter t; $\bar{y}_t$ = average paid claim costs for the year ending in quarter t; t = time, in quarters; and w_t = private sector wages per man-hour in quarter t.

Note: The estimation period for each time trend equation is the twelve quarters preceding the first forecast period. For each linear econometric equation, the estimation period begins in 1964.1 and ends in the quarter preceding the first forecast period. The numbers in parentheses below the estimated coefficients are t-statistics.

[a] Numbers in this column are Durbin-Watson d statistics. Rejection of the hypothesis of no serial correlation (at the 5 percent level) is indicated by an asterisk. The test for 1973.3 is inconclusive.

[b] This column presents the GOLDFELD-QUANDT (1965) test statistics for heteroskedasticity. The hypothesis of homoskedasticity is rejected if the statistic value exceeds F_{α, ν_1, ν_2}. The degrees of freedom parameters, ν_1 and ν_2, are given in parentheses below the statistics. At the 5 percent level the hypothesis is rejected for each of the last four equations (1974.3 through 1977.3).

TABLE 5: AUTOREGRESSIVE EQUATIONS FOR AUTOMOBILE BODILY INJURY LIABILITY
AVERAGE PAID CLAIM COSTS:
EQUATIONS IN LOWER PANEL ADJUSTED FOR HETEROSKEDASTICITY

Equation for

$$y_t = \hat\rho y_{t-1} + \hat\alpha(1-\hat\rho) + \hat\beta(w_t - \hat\rho w_{t-1}) + \hat\gamma_1 S_1 + \hat\gamma_2 S_2 + \hat\gamma_3 S_3$$

Forecast Beginning in	$\hat\alpha$	$\hat\beta$	$\hat\gamma_1$	$\hat\gamma_2$	$\hat\gamma_3$	$\hat\rho$	R^2
1971.3	−110.946 (−1.559)	344.294 (17.381)	−41.1317 (−3.037)	N.E.	−74.0772 (−4.774)	.440689 (2.145)	.9763
1972.3	N.E.	313.435 (41.365)	−40.3019 (−2.773)	N.E.	−70.9714 (−4.473)	.527511 (3.129)	.9711
1973.3	N.E.	313.612 (39.736)	−41.3901 (−3.070)	N.E.	−70.3248 (−5.015)	.586522 (4.208)	.9767
1974.3	127.914 (1.556)	284.169 (16.527)	−38.1477 (−2.902)	N.E.	−81.546 (−5.908)	.620412 (5.015)	.9807
1975.3	288.274 (2.435)	252.929 (12.912)	−45.5498 (−3.449)	N.E.	−90.9309 (−6.458)	.712105 (6.303)	.9822
1976.3	235.516 (3.021)	260.034 (19.667)	−46.9951 (−3.621)	N.E.	−83.5657 (−6.206)	.654109 (5.897)	.9862
1977.3	221.158 (3.538)	263.160 (25.981)	−49.726 (−4.069)	N.E.	−84.430 (−6.732)	.638744 (6.030)	.9901
1971.3	−142.405 (−2.253)	353.898 (18.752)	−38.1636 (−3.239)	N.E.	−79.2642 (−6.201)	.414891 (2.464)	.966
1972.3	N.E.	317.833 (35.971)	−42.1668 (−3.578)	N.E.	−84.5401 (−7.021)	.56165 (4.371)	.969
1973.3	N.E.	317.593 (37.473)	−42.2427 (−3.789)	N.E.	−83.5804 (−7.436)	.57197 (4.815)	.978
1974.3	106.858 (1.252)	290.600 (15.467)	−40.8001 (−3.276)	N.E.	−82.9638 (−6.396)	.63102 (5.169)	.872
1975.3	234.868 (1.971)	264.464 (11.848)	−45.5516 (−3.6655)	N.E.	−89.5063 (−6.896)	.71128 (6.390)	.852
1976.3	231.793 (2.849)	261.226 (18.589)	−46.6890 (−3.645)	N.E.	−84.1028 (−6.343)	.66127 (6.016)	.980
1977.3	220.898 (3.811)	262.653 (29.542)	−51.6890 (−4.190)	N.E.	−83.8710 (−6.609)	.61937 (5.717)	.995

Key: y_t = average paid claim costs in quarter t; w_t = private sector wages per man-hour in quarter t; and S_i = 1.0 in quarter i of each year, 0 otherwise.

Note: The estimation period for each equation begins in 1964.1 and ends in the quarter preceding the first forecast period. The number in parentheses below each estimated coefficient is the ratio of the coefficient to its estimated standard deviation. N.E. = found to be insignificant in earlier versions of the equations and thus omitted from the final version.

TABLE 6: RESULTS OF EIGHT QUARTER FORECASTS OF AUTOMOBILE BODILY INJURY LIABILITY AVERAGE PAID CLAIM COSTS

First Period of Forecast	Total Predicted Change Errors [a]				Mean Absolute Percent. Errors [a]				Root Mean Square Percent. Errors [a]				Shift Index [b]
	ET	OLS	AR	ARH	ET	OLS	AR	ARH	ET	OLS	AR	ARH	
1971.3	10.1%	3.4%	5.6%	7.0%	2.6%	3.7%	4.3%	5.2%	2.8%	4.5%	5.0%	6.0%	63%
1972.3	-8.4	0.1	0.4	1.6	5.9	4.6	2.4	2.7	6.0	5.7	3.4	3.6	90
1973.3	-4.6	3.2	6.1	7.6	2.2	4.5	4.3	5.1	2.6	5.4	5.0	5.7	126
1974.3	-4.6	5.3	6.3	7.3	6.8	5.3	4.6	5.3	7.2	5.6	5.1	5.6	161
1975.3	-5.0	0.7	-0.4	1.4	2.2	2.4	1.5	1.9	2.4	2.8	2.0	2.2	171
1976.3	2.4	2.5	2.8	2.9	0.5	2.0	1.2	1.2	0.7	2.4	1.6	1.6	85
1977.3 [c]	3.9	1.5	2.6	2.5	0.8	1.8	1.7	1.6	1.1	2.2	2.0	2.0	80
Absolute Averages	5.6%	2.4%	3.5%	4.3%	3.0%	3.5%	2.9%	3.3%	3.3%	4.1%	3.4%	3.8%	

Key: ET = exponential trend model, OLS = ordinary least squares model, AR = autoregressive model without heteroskedasticity adjustment, ARH = autoregressive model with heteroskedasticity adjustment.

Note: The econometric forecasts (columns labelled OLS, AR, and ARH) are based on predicted values of the private sector wage rate promulgated by Wharton Econometric Forecasting Associates, Inc. The equations were estimated using revised values of the wage rate series, which differ slightly in most cases from the published values available at the time the Wharton forecasts were made. Thus, the predicted value of the wage rate $\hat{w}_t$, for use in the claim cost forecasts is $\hat{w}_t = w_{0c} \cdot \hat{w}_{tw}/w_{0p}$, where w_{0c} = the current (revised) value of w for the period preceding the first forecast period, w_{0p} = the (preliminary) value of w for the period preceding the first forecast period at the time the Wharton forecast was made, and $\hat{w}_{tw}$ = the Wharton forecast of w for period t.

[a] Total predicted change error (TPCE), mean absolute percentage error (MAPE), and root mean square percentage error (RMSPE) are defined in equations (13), (14), and (15) in the text.

[b] The shift index is computed as follows: $I = [(y_8/y_0)^{1/8} - 1] \div [(y_0/y_{-12})^{1/12} - 1]$, where y_8 = actual average paid claim costs in the last quarter of the forecast period, y_0 = actual claim costs in the quarter preceding the forecast period, and y_{-12} = actual claim costs thirteen quarters prior to the forecast period.

[c] This forecast period runs from 1977.3 through 1978.2.

with adjustment for autocorrelation and for both autocorrelation and heteroskedasticity, using Method 2 [15]. These results are shown in Table 5.

The forecasting errors for the four sets of bodily injury equations are presented in Table 6. Using the TPCE as the measure of accuracy, the OLS equations give the best results in five of the seven forecast periods and the exponential trend model is superior in two. In one of these periods, however, the trend line underpredicts while the econometric models overpredict. In the other, the trend line is better than the OLS equation by only 0.1 percentage point. When MAPE and RMSPE are used to gauge forecasting accuracy, the exponential trend equation is superior to the econometric equations in four of the seven periods. However, two of the periods when at least one of the econometric equations is better are those with the highest unanticipated inflation.

6. SUMMARY AND CONCLUSIONS

The results of this study indicate that econometric models can be used to forecast automobile insurance paid claim costs more accurately than the exponential trend models now in use in the United States. The superiority of the econometric models is especially evident during periods of high unanticipated inflation. Although adjustments for seasonality, autocorrelation, and heteroskedasticity often improve forecasting accuracy, even the OLS models perform quite well. Adoption of the econometric approach may enable insurance companies to eliminate some of the uncertainty associated with claims cost inflation.

REFERENCES

COOK, C. (1970). Trend and Loss Development Factors, *Proceedings of the Casualty Actuarial Society*, **52**, 1-14.

DURBIN, J. (May 1970). Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables, *Econometrica*, **38**, no. 3, 410-421.

GLEJSER, H. (1969). A New Test for Heteroskedasticity, *Journal of the American Statistical Association*, **64**, 316-323.

GOLDFELD, S. M. and R. E. QUANDT. (1965). Some Tests for Homoskedasticity, *Journal of the American Statistical Association*, **60**, 539-547.

HALL, B., HALL, R., and S. BECKETTI. (1977). Time Series Processor, Version 3.5, Privately published. Stanford, Calif.

HARRISON, M. J. and B. P. M. McCABE. (1979). A Test for Heteroskedasticity Based on Ordinary Least Squares Residuals, *Journal of the American Statistical Association*, **74**, 494-499.

HARVEY, A. C. and G. D. A. PHILLIPS. (1974). A Comparison of the Power of Some Tests for Heteroskedasticity in the General Linear Model, *Journal of Econometrics*, **2**, 307-316.

KLEIN, L. R. and E. BURMEISTER. (1976). *Econometric Model Performance: Comparative Simulation Studies of the U.S. Economy*, University of Pennsylvania Press, Philadelphia.

KMENTA, J. (1971). *Elements of Econometrics*, The Macmillan Co., New York.

SPENCER, B. G. (1975). The Small Sample Bias of Durbin's Test for Serial Correlation, *Journal of Econometrics*, **3**, 249-254.

WAELBROECK, J., ed. (1976). *The Models of Project LINK*, North-Holland Publishing Co., Amsterdam.

[15] For the first three sample periods, $\hat{P}(x_t) = \hat{\gamma}_1 x_t^2$; while for the last four periods $\hat{P}(x_t) = \hat{\gamma}_0 + \hat{\gamma}_1 x_t^2$.

APPENDIX: PRINCIPAL DATA SERIES

Quarter	PD	WR	Quarter	PD	BI	WR	Quarter	PD	BI	WR
1954.1	1.0	1.95	1962.3	1.4078		2.76	1971.1	2.7864	1.5276	4.397
1954.2	0.9903	1.96	1962.4	1.4466		2.79	1971.2	2.9029	1.7074	4.468
1954.3	0.9709	1.98	1963.1	1.4272		2.82	1971.3	2.9515	1.6273	4.540
1954.4	1.0097	2.0	1963.2	1.4660		2.84	1971.4	3.0291	1.6592	4.570
1955.1	1.0097	2.01	1963.3	1.4757		2.86	1972.1	2.9417	1.6228	4.679
1955.2	1.0291	2.03	1963.4	1.5146		2.90	1972.2	2.9320	1.5781	4.737
1955.3	1.0194	2.06	1964.1	1.4854	1.0	2.947	1972.3	3.0194	1.5852	4.797
1955.4	1.0680	2.08	1964.2	1.5437	1.0341	2.972	1972.4	3.1165	1.6804	4.858
1956.1	1.0680	2.12	1964.3	1.5534	0.9330	3.010	1973.1	3.0874	1.7004	5.006
1956.2	1.0777	2.15	1964.4	1.6214	1.0364	3.026	1973.2	3.1845	1.7474	5.090
1956.3	1.0777	2.19	1965.1	1.6117	1.04	3.040	1973.3	3.1845	1.6439	5.179
1956.4	1.1456	2.22	1965.2	1.6796	1.0811	3.067	1973.4	3.3495	1.7709	5.272
1957.1	1.1553	2.25	1965.3	1.6990	1.0129	3.106	1974.1	3.2330	1.8696	5.384
1957.2	1.1845	2.28	1965.4	1.7961	1.0999	3.141	1974.2	3.2621	1.9389	5.538
1957.3	1.1845	2.31	1966.1	1.7864	1.0764	3.185	1974.3	3.4078	1.8284	5.675
1957.4	1.2427	2.33	1966.2	1.8447	1.1340	3.241	1974.4	3.6117	1.9130	5.837
1958.1	1.2330	2.35	1966.3	1.8447	1.1011	3.292	1975.1	3.6796	1.8848	5.999
1958.2	1.2427	2.37	1966.4	1.9515	1.1915	3.331	1975.2	3.7476	1.9965	6.097
1958.3	1.2427	2.40	1967.1	1.9417	1.2045	3.384	1975.3	3.8058	2.0364	6.191
1958.4	1.2718	2.42	1967.2	1.9903	1.2468	3.433	1975.4	3.9417	2.1633	6.305
1959.1	1.2524	2.44	1967.3	2.0291	1.2233	3.472	1976.1	3.9806	2.1293	6.439
1959.2	1.2913	2.46	1967.4	2.1068	1.2667	3.515	1976.2	4.0874	2.1774	6.578
1959.3	1.2816	2.48	1968.1	2.1262	1.2785	3.609	1976.3	4.1845	2.1598	6.707
1959.4	1.3398	2.51	1968.2	2.2136	1.3619	3.669	1976.4	4.4175	2.2996	6.842
1960.1	1.3204	2.55	1968.3	2.2330	1.2808	3.721	1977.1	4.4078	2.2985	7.007
1960.2	1.3204	2.58	1968.4	2.3592	1.3690	3.789	1977.2	4.5922	2.3925	7.118
1960.3	1.3301	2.59	1969.1	2.3592	1.3596	3.842	1977.3	4.7573	2.3678	7.240
1960.4	1.3689	2.62	1969.2	2.4466	1.4101	3.904	1977.4	4.9612	2.4618	7.355
1961.1	1.3010	2.63	1969.3	2.4951	1.4148	3.966	1978.1	4.8835	2.3960	7.613
1961.2	1.3495	2.65	1969.4	2.6117	1.4160	4.035	1978.2	5.1456	2.4982	7.758
1961.3	1.3592	2.67	1970.1	2.5631	1.4348	4.098				
1961.4	1.4078	2.69	1970.2	2.6796	1.5018	4.168				
1962.1	1.3689	2.73	1970.3	2.7379	1.4606	4.253				
1962.2	1.3981	2.75	1970.4	2.8350	1.5018	4.306				

Key: *BI* and *PD* = automobile bodily injury and property damage liability paid claim costs, respectively. These series are expressed as indices with the earliest period = 1.0. *WR* = the private sector wage rate.