Forecasting Workers Compensation Severities and Frequency Using the Kalman Filter

Jonathan Evans and Frank Schmid

Motivation. Estimating trend rates of growth of severity and frequency is crucial to workers compensation ratemaking. Such trend rates can be estimated using unobserved components models and structural time series models. These two types of models derive from parsimonious and transparent data-generating processes and, in the case of structural time series models, allow the researcher to incorporate economically meaningful explanatory variables into a time series framework. When specified in state-space form, unobserved components models and structural time series models become available to the Kalman filter estimation technique. The Kalman filter explicitly accounts for possible measurement errors in the observed severity and frequency data.
Model. Structural time series models, which nest unobserved components models, are applied to state-level time series data for (on-leveled and wage-adjusted) indemnity and medical severities, and for frequency. Parameter estimates, hypothesis tests, and growth forecasts are provided for by the software package STAMP. STAMP is especially designed for estimating unobserved components and structural time series models.
Results. NCCI developed a production process that employs unobserved components and structural time series models to state-level data of indemnity and medical severities, and frequency. Trend growth forecasts generated with such models were presented in state advisory forums and served as a consideration in rate filings.
Conclusions. NCCI's experience with Kalman-filtered estimation of trend rates during the policy year 2006 rate-filing season was encouraging. NCCI anticipates continued use of unobserved components models and structural time series models in future rate filings.
Availability. STAMP is an easy-to-use windows-driven software package that runs on the GiveWin platform. STAMP and GiveWin are available from Timberlake Consultants Ltd.
Keywords. Workers compensation, trend growth rates, Kalman filter, unobserved components model, structural time series models, state-space modeling

1. INTRODUCTION

Forecasting frequency and severity is crucial to workers compensation ratemaking. Such forecasting is performed using time series models, which are models that account for the time dependence in the observed data. In many time series models, this time dependence is modeled as a (potentially rather complex) autoregressive structure, as is the case in ARIMA (Auto-Regressive Integrated Moving Average) or ARMA (Auto-Regressive Moving Average) models. To many, such autoregressive structures appear mechanistic. In search for more transparent and parsimonious representations of the underlying data-generating processes, unobserved components (UC) models and, as an extension to UC models, structural time series (STS) models have been developed. In UC models the quantities of most interest are not directly observed and must be estimated using both empirical data and estimates of underlying statistical parameters (sometimes called *hyperparameters* in this context). STS models are linear combinations of UC models for the time series of interest and standard linear regression models including explanatory variables that are exogenous to the time series of interest. For example an STS model for stock market prices might combine a UC random walk with drift for the logarithm of stock prices and a standard linear regression for changes in the logarithm of stock market prices against recent changes in interest rates.

Forecasting is a signal extraction and signal extrapolation exercise. Signal extraction is the process of filtering out measurement errors from empirical data. Measurement errors include the total impact from all sources of *noise*, deviations of the empirical data from the underlying signal that do not affect the expected values of future observations (such as medical or indemnity severities, and frequency). In forecasting, the signal is the quantity of interest, because it is the signal that determines the expected values of future observations. Specifically, it is the objective of a forecasting model to elicit from historical observations the process that generates the unobservable signal. Because the forecasting model replicates the data-generating process of the signal (instead of fitting historical observations), the quality of these models cannot be judged by the (in-sample) fit to the observed data, as gauged, for instance, by the R^2. In fact, good fit to heretofore observed data harbors the risk of overfitting. Such overfitting implies that the (in-sample) fits and (out-of-sample) forecasts may not center on the signal, thus giving rise to potentially large forecasting errors.

As an example, consider a game of dice, where each die has six faces, the number of spots ranging from 1 to 6. In any toss of a pair of dice, the expected value of the outcome is 7. This expected value is the signal, which manifests itself as the mean outcome as the number of tosses goes to infinity. The difference between the observations and the signal is noise. The signal offers an unbiased forecast for any future toss. Thus, among all possible forecasting models, the one that simply produces this time-invariant signal as its forecast has the lowest expected root mean squared error. Yet, this model offers the worst in-sample fit possible, as the model has no explanatory power with regards to the variation of the outcome around the expected value. Not surprisingly, a least-squares regression of the 36 possible outcomes on the time-invariant signal reveals an R^2 equal to zero.

The risk of overfitting awards parsimony a critical role in time series modeling. UC models are conducive to such parsimonious modeling as the underlying data-generating process is highly transparent. UC models, and their extension, STS models, can be written in state-space form (defined in section 2.2), which makes these models available to the Kalman filter estimation technique. The Kalman filter has been developed in engineering as a signal extraction algorithm and, as such, recommends itself for estimating forecasting models. In fact, the Kalman filter is an estimation technique that explicitly accounts for possible measurement errors in the reported data.

NCCI estimates UC and STS models using the software packages STAMP and SsfPack. SsfPack is a collection of functions for state-space modeling, including maximum likelihood-estimation (MLE) and Kalman-filtered estimation and smoothing. This package runs in two alternative environments: the programming language Ox (Koopman et al. [11]) and the platform S-Plus (Zivot et al.[17]). We use SsfPack within Ox Professional on the GiveWin platform. STAMP (Koopman et al. [10]) also runs on the GiveWin platform. Ox Professional, GiveWin, and STAMP are distributed by Timberlake Consultants Ltd. S-Plus is a commercial platform available from Insightful Corporation. The models presented here were estimated using STAMP. Due to the complexity of code development, practical implementation of the Kalman filter in actuarial applications generally requires the acquisition of a preexisting specialized statistical software package from an external vendor.

NCCI developed a production process that employs unobserved components and structural time series models for indemnity and medical severities and for frequency for more than 30 U.S. states. Trend growth forecasts derived from these models were presented in state advisory forums and served as a consideration in rate filings.

1.1 Research Context

The material in this paper falls under CAS Research Categories II.G.12 Actuarial Applications and Methodologies/Ratemaking/Trend and Loss Development and III.H.15 Financial and Statistical Models/Statistical Models and Models/Time Series. Econometric models for actuarial trends have been dealt with in Hartwig et al.[5], Lommele and Sturgis.[12], McGuinness[13], and Van Slyke[15]. Credibility adjusted trending has been discussed in Venter[16]. None of these sources utilize the Kalman filter.

1.2 Objective

Economic support for actuarial trending of workers compensation losses at NCCI currently includes UC and STS models for forecasting (on-leveled and wage-adjusted) medical and indemnity severities, and frequency (number of claims, divided by on-leveled and wage-adjusted premium). (Wage-adjusting brings past exposure, as gauged by payroll, up to current wage levels; on-leveling brings past loss experience up to current benefit levels.) This paper describes current practice at NCCI of estimating such models using the Kalman filter. In addition to this set of three single-equation models, NCCI operates a Bayesian five-equation state-space forecasting model for severities, frequency, and the corresponding loss ratios—this multi-equation model, which accounts for add-up constraints and contemporaneous (cross-equation) covariances, is estimated using the Metropolis-Hastings algorithm. This paper is written from the perspective of actuarial researchers using preexisting statistical software packages from external vendors and does not include algorithmic details for statistical methods.

1.3 Outline

In Section 2.1, we describe the data-generating processes that underlie UC and STS models, and we put these models in state-space form. In section 2.2, we discuss the Kalman filter estimation technique and show how ML estimates for the moments are obtained from the Kalman filter output. The authors caution that readers need not completely understand the formulaic details in section 2.2 to understand the rest of this paper. Section 3 describes an implementation of UC and STS models in indemnity and medical severities and frequency forecasting. Section 4 concludes.

2. BACKGROUND AND MODELS

2.1 Unobserved Components and Structural Time Series Models

STS models are linear combinations of UC models and standard linear regression models. We start out by describing the data-generating processes of UC models and then expand these models to the STS framework.

UC models derive from the concept of Gaussian *innovations*, as exemplified in Brownian motion. Unlike noise, innovations propagate forward in time and affect the expected values of future observations. In their most basic (and, hence, most restrictive) form, these models postulate that innovations to the (unobserved) signal of a given (observable) variable are draws from the normal distribution. Put differently, the signal in question follows a random walk. Let y_t be the variable and θ_t the signal, then we can write the *local level model* in Equations 2.1.1 through 2.1.3 as follows:

$$y_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\sigma_\varepsilon^2),\ \ t = 1,\dots,T \tag{2.1.1}$$

$$\theta_t = \mu_t \tag{2.1.2}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2) \qquad (2.1.3)$$

The local level model is written in state-space form. The variable μ_t is the only state variable and, by definition, unobservable. The variable μ_t describes the time t *level* of the unobservable signal θ_t and is subject to the Guassian innovation η_t. The observed dependent variable y_t is the sum of this signal and Gaussian noise ε_t. Inserting Equation (2.1.2) into Equation (2.1.1) delivers the *measurement equation.* Equation (2.1.3) is the *transition equation*, which describes the trajectory of the state variable μ_t.

Local level models apply when the signal follows a random walk. Variables that follow random walks exhibit high degrees of persistence, as all innovations are permanent. An example of such a highly persistent variable is the rate of CPI (Consumer Price Index) inflation (see, for instance, Koopman et al.[11] and Green[4]).

Signals may exhibit drift. If this drift is stochastic, we obtain the *local linear model*, which is described in Equations 2.1.4 through 2.1.7:

$$y_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \ t = 1, ..., T \qquad (2.1.4)$$

$$\theta_t = \mu_t \qquad (2.1.5)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2) \qquad (2.1.6)$$

$$\beta_t = \beta_{t-1} + \zeta_t, \qquad \zeta_t \sim N(0, \sigma_\zeta^2) \qquad (2.1.7)$$

The state variable μ_t indicates the *level* of the signal, and the state variable β_t describes the *slope* (or, synonymously, drift) of the signal. As with the level, the slope is governed by a Gaussian, permanent innovation ζ_t. Because there are two state variables in the local linear model, there are two transition equations, which are Equations 2.1.6 and 2.1.7.

An example of a variable the trajectory of which may be described using a local linear model is the logarithmic stock market total-return index, where β_t indicates the expected log return (or, equivalently, the drift in the logarithmic stock price). For $\sigma_\zeta^2 = 0$, the slope is non-stochastic. In the stock market example, non-stochastic drift implies constant expected return.

A third model of interest follows from the local level model by means of integration. Assume that the CPI rate of inflation indeed describes a random walk, as empirical studies indicate, and measure the rate of inflation by the first difference in the logarithmic price level. In this case then, the logarithmic price level follows an integrated random walk. For the ***integrated random walk*** (which sometimes is called ***smooth trend***) model, we can write in Equations 2.1.8 through 2.1.11:

$$y_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\sigma_\varepsilon^2),\ t = 1,\ldots,T \tag{2.1.8}$$

$$\theta_t = \mu_t \tag{2.1.9}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1}, \tag{2.1.10}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \qquad \zeta_t \sim N(0,\sigma_\zeta^2) \tag{2.1.11}$$

The integrated random walk model results from the local linear model for $\sigma_\eta^2 = 0$.

The described types of UC models rest on parsimonious data-generating processes, which makes them appealing for signal-extraction purposes. On the other hand, these models are not cognizant of economic, causal relations that may exist between the dependent variable in question, y_t, and a vector of variables of economic activity, $(x_{1,t}, x_{2,t}, \ldots, x_{n,t})$. UC models can be expanded to STS models by adding a standard regression component, thus enabling such models to account for pertinent economic relations. When expanding the most general UC model—the local linear model—to an STS model, we can write in Equations 2.1.12 through 2.1.16:

$$y_t = \theta_t + \gamma_t \cdot x_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0,\sigma_\varepsilon^2) \tag{2.1.12}$$

$$\theta_t = \mu_t \tag{2.1.13}$$

$$\gamma_t = \gamma_{t-1} + \upsilon_t, \qquad \upsilon_t \sim N(0,\sigma_\upsilon^2) \tag{2.1.14}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma^2_\eta) \tag{2.1.15}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim N(0, \sigma^2_\zeta) \tag{2.1.16}$$

The STS model above has one explanatory variable, x_t. The regression parameter of this variable, γ_t, follows a random walk. Alternatively, this parameter may be specified as stationary ($\gamma_t = \bar{\gamma} + \upsilon_t$) or time-invariant ($\sigma^2_\upsilon = 0$). At NCCI, when employing time-variant parameters, we estimate such STS models using SsfPack. The software package STAMP (we use version 6.21) can handle currently only time-invariant parameters. For short data series, time invariance in γ_t is an appropriate constraint to avoid over-parametrization. Such time-invariance presumes that the variable x_t is measured without error and that the economic relation depicted in the above measurement equation is time-invariant—these are standard assumptions in ordinary linear regression models.

In the next section, we describe the Kalman filter technique for estimating the state variables, and the accompanying ML estimation of the moments. Further, we discuss the relation between the Kalman filter and the Bühlmann credibility criterion.

2.2 The Kalman Filter

The Kalman filter was invented in 1960 by Rudolf Kalman (Kalman [7]) and saw almost immediate application in real-time signal processing for spacecraft. Up to the present, the Kalman filter is widely used in various aspects of aerospace operations, such as radar. The filter acts on an observed time series by removing an estimate of measurement noise. Thus, the filtered series represents an estimate for the underlying process of the signal, that is, the observed variable, purged of noise. The Kalman filter introduces into time series modeling the fundamental statistical philosophy that real-world observations are only shadows of ideal Platonic forms (Plato [14]).

The Kalman filter works in the context of time series models expressed in state-space form. The state-space form specifies a transition vector equation (Equation 2.2.1) for unobserved state variables of interest, and an associated measurement vector equation (Equation 2.2.2) for the observed series (Harvey[6] and Durbin and Koopman[3]). The transition equations describe the transition of the state variables from state t to state $t+1$. The measurement equations describe the relations between the signals and the state variables and, at the same time, account for measurement noise as the difference between the observed variables and the respective signals.

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t,\ \mathrm{E}(\eta_t) = 0,\ \mathrm{Var}(\eta_t) = Q_t \tag{2.2.1}$$

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \mathrm{E}(\varepsilon_t) = 0,\ \mathrm{Var}(\varepsilon_t) = H_t \tag{2.2.2}$$

The vector α_t contains the unobservable state variables. The matrices T_t, R_t, and Z_t, as well as the parameter vectors c_t, and d_t are assumed to be non-stochastic; typically, these variables and parameters are known and may (but need not) be time-invariant. In engineering applications, the matrices of the variances of the innovations and the measurement errors, Q_t and H_t, respectively, are often determined by actual physical calibration with instruments. In financial and economic analyses, these moments are estimated using the ML approach. This ML estimation can easily be obtained through a decomposition of the prediction error of the Kalman filter (Kim and Nelson[8]). The Kalman filter is presented in Equations 2.2.3 through 2.2.7.

$$a_{t|t-1} = T_t a_{t-1} + c_t \tag{2.2.3}$$

$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t' \tag{2.2.4}$$

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t|t-1} - d_t) \tag{2.2.5}$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \tag{2.2.6}$$

$$F_t = Z_t P_{t|t-1} Z_t' + H_t \tag{2.2.7}$$

For initial values, it is assumed that $P_{2|1}$ is very large and that $a_{2|1} = 0$.

The coefficients $a_{t|t-1}$ and a_t represent estimates for α_t before and after y_t is observed, respectively.

The exists an analogy between the Kalman filter and the Bühlmann credibility criterion (Venter[18]); to make the analogy more apparent, assume $Z_t = 1$ and $d_t = 0$. Equation 2.2.5 contains the Bühlmann credibility-like term:

$$\frac{P_{t|t-1}}{P_{t|t-1} + H_t}$$

If y_t , $a_{t|t-1}$, and a_t are interpreted as the indication, the complement of credibility, and the credibility-weighted estimate, respectively, then Equation 2.5 is effectively a Bühlmann credibility estimate since $P_{t|t-1}$ and H_t can be interpreted as estimates of the variances of $a_{t|t-1}$ and y_t, respectively.

Note that the Kalman filter only estimates the series of underlying states α_t, given the observed series y_t and assumed values for the variance parameters contained in Q_t and H_t. These variance parameters must still be estimated via ML. In general, likelihood functions for time series models based on prior estimates of observations conditional on all previous observations can be stated as in Equation 2.2.8, where θ represents the parameter values:

$$L(y;\theta)=\prod_{t=1}^{T} p(y_t \mid y_{t-1},\dots,y_1) \tag{2.2.8}$$

The Kalman filter estimates can be used to derive prior means and variances of not yet observed points, conditional on the previous observations. Since the actual observations the are conditionally normally distributed, the log-likelihood function can be written as Equation 2.2.9, where N is the number of scalar components of y_t:

$$l(y;\theta)=-\frac{NT}{2}\log(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\log|F_t|-v_t{}'F_t^{-1}v_t \tag{2.2.9}$$

$$v_t = y_t - Z_t a_t - d_t, \tag{2.2.10}$$

The Kalman filter works reasonably well even for some time series shorter than 30 points, although the filtered series may behave erratically on the first few data points.

The filtered estimates are predictions for the time-t vector of state variables, based on information available at time $t-1$. In the context of economics, these filtered estimates may be interpreted as expectations for time t that economic agents formed based on the information available at time $t-1$. Thus, the time-t filtered estimates of the rate of inflation may serve as a gauge of inflation expectations (Koopman et al. [10]).

Typically, the researcher is not interested in time-t estimates for the state vector that are based on information available only at time $t-1$. As in standard regression analysis, the researcher looks for time-t estimates of the state vector that use all information available as at the end of the data series, that is, as at time $T \geq t$. Such estimates can be obtained through a backward moment-smoothing algorithm (Bryson and Ho [1], de Jong [2], Kohn and Ansley [9]). This algorithm is presented in Equations 2.2.11 through 2.2.16, where the initial conditions are set as $N_T = r_T = 0$:

$$K_t = T_t P_{t|t-1} Z_t' F_t^{-1} \tag{2.2.11}$$

$$e_t = F_t^{-1} v_t - K_t' r_t \tag{2.2.12}$$

$$r_{t-1} = Z_t' e_t + T_t' r_t \tag{2.2.13}$$

$$D_t = F_t^{-1} + K_t' N_t K_t \tag{2.2.14}$$

$$L_t = T_t - K_t Z_t \tag{2.2.15}$$

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t \tag{2.2.16}$$

Equations 2.2.17 through 2.2.20 present the moment-smoothed estimates of the stochastic elements and their associated variances:

$$\mathrm{E}[\varepsilon_t \mid \{y_1, \ldots, y_T\}] = H_t e_t \tag{2.2.17}$$

$$\mathrm{Var}[\varepsilon_t \mid \{y_1, \ldots, y_T\}] = H_t D_t H_t \tag{2.2.18}$$

$$\mathrm{E}[\eta_t \mid \{y_1, \ldots, y_T\}] = R_t' Q_t R_t r_t \tag{2.2.19}$$

$$\mathrm{Var}[\eta_t \mid \{y_1, \ldots, y_T\}] = R_t' Q_t R_t N_t R_t' Q_t R_t \tag{2.2.20}$$

3. IMPLEMENTATION

We now demonstrate how to apply UC and STS models to state-level series of (on-leveled and wage-adjusted) indemnity and medical severities and of frequency (number of claims, divided by on-leveled and wage-adjusted premium). The objective is to forecast the growth factor $1+g_{T,T+3}$ for the indemnity and medical severities that applies to the 3-year period between the last observed period T and the future period $T+3$. (The number of years may not be an integer; for instance, the time interval may range from T to $T+3+\varepsilon$, $0<\varepsilon<1$, in which case the applicable growth factor reads $1+g_{T,T+3+\varepsilon}$.)

There are several routes to arriving at such a 3-year growth factor. One approach is to estimate directly 3-year rates of growth $\hat{g}_{T,T+3}$ from (successive and non-overlapping) 3-year time periods. This method requires long data series (as the number of data points is, at maximum, one-third of the number of annual observations) and, hence, is not an option at NCCI. An alternative route is to estimate annual rates of growth and then tally up the annual forecasts for the time periods $T+1$, $T+2$, and $T+3$ in order to obtain the 3-year rate of growth from T to $T+3$. Tallying up forecast rates of growth is not straightforward as these forecasts are random variables and annual compounding involves nonlinear transformations. For instance, let $\hat{g}_{T+1}$, $\hat{g}_{T+2}$, and $\hat{g}_{T+3}$ be the forecasts for the annual rates of growth and calculate the forecast for the 3-year growth rate by means of compounding: $\hat{g}_{T,T+3}=(1+\hat{g}_{T+1})\cdot(1+\hat{g}_{T+2})\cdot(1+\hat{g}_{T+3})-1$. In this case then, if the 3 annual forecasts $\hat{g}_{T+i}$ are unbiased forecasts of the actual annual rates of growth g_{T+f} $(f=1,2,3)$, $\hat{g}_{T,T+3}$ is a biased forecast for the pertinent future 3-year rate of growth $g_{T,T+3}=(1+g_{T+1})\cdot(1+g_{T+2})\cdot(1+g_{T+3})-1$.

We arrive at our forecast for the growth rate $g_{T,T+3}$ by means of estimating and tallying up logarithmic rates of growth—that is, first differences in natural logarithms. We choose this approach because, here, our interest is to estimate the geometric mean of the (continuously compounded) annual rates of growth rather than the arithmetic mean. Logarithmic rates of growth are additive; thus we can write: $\hat{g}^{\log}_{T,T+3}=\hat{g}^{\log}_{T+1}+\hat{g}^{\log}_{T+2}+\hat{g}^{\log}_{T+3}$. (When there is the fraction ε of an incomplete fourth year, then the multi-year growth rate amounts to $\hat{g}^{\log}_{T,T+3+\varepsilon}=\hat{g}^{\log}_{T+1}+\hat{g}^{\log}_{T+2}+\hat{g}^{\log}_{T+3}+\varepsilon\cdot\hat{g}^{\log}_{T+4}$.) This additivity property implies that the sum of the annual forecast growth rates is indeed an unbiased estimator of the multi-year logarithmic rate of growth. By means of invoking normality, it is possible to calculate a standard error for $\hat{g}^{\log}_{T,T+3}$ from the variances of the annual growth rates $\hat{g}^{\log}_{T+1}$, $\hat{g}^{\log}_{T+2}$, and $\hat{g}^{\log}_{T+3}$. These standard errors then enable us to compute confidence bounds around $\hat{g}^{\log}_{T,T+3}$. This provides valuable information, rarely if ever available from traditional actuarial trend analyses, about the uncertainty of trend estimates.

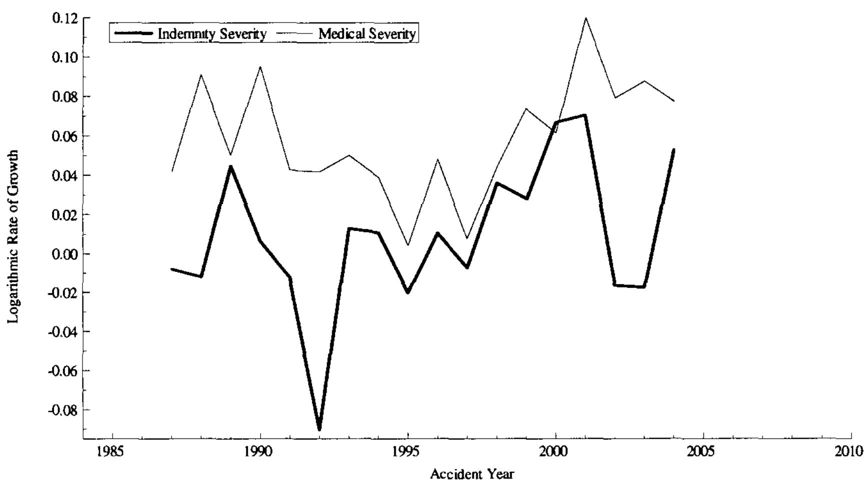
The pertinent severity series are on a "paid" basis. The severity and frequency data are from an anonymous U.S. state and refer to the policy year 2006 rate-filing season. These data series range from 1986 through 2004, thus affording 18 annual growth rate observations. The model estimates presented below are for illustration purposes only and are not necessarily identical to the estimates used in the rate-filing for the anonymous state in question.

According to the NBER (National Bureau of Economic Research, www.nber.org), there are two economic recessions that fall into the analyzed 1986-2004 period. Both of these recessions lasted for 8 months, as measured from peak to trough. The 1990/91 recession lasted from July 1990 to March 1991, and the 2000 recession lasted from March to November. This fluctuation in economic activity is potentially important for frequency. For instance, it can be shown that the growth rate of BLS (Bureau of Labor Statistics, www.bls.gov) on-the-job injury rates correlate with the change in the rate of unemployment. Similarly, it is common for NCCI states that the growth rate of frequency correlates with the change in the state-level rate of unemployment.

Chart 1 shows for the anonymous state in question the log growth rates for the on-leveled and wage-adjusted indemnity and medical severities. Chart 2 exhibits the log growth rate of frequency, along with the first difference in the percentage rate of unemployment (which has been divided by 10 in this exhibition, for scaling purposes).

Chart 1

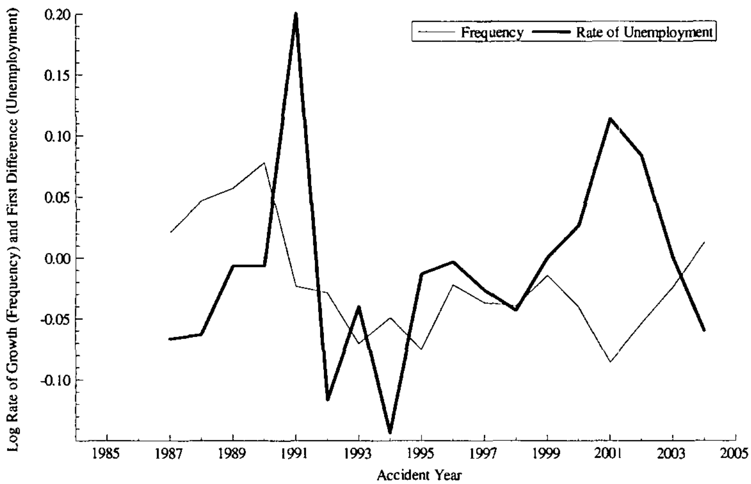
Logarithmic Growth Rates of Indemnity and Medical Severities, State-Level Data, Accident Years 1987-2004



Visual inspection of Charts 1 and 2 suggests that the rates of growth of the indemnity and medical severities and of frequency follow random walks—thus, the local level model applies. A *Unit-root process* includes an autoregressive AR(1) coefficient of unity. We do not employ unit-root tests, such as Dickey-Fuller (see, for instance, Greene [4]). This is because, for short time series, these tests have little power, that is, the test are deficient in their ability to reject the null hypothesis of the presence of a unit root. What complicates matters for the frequency series is that this variable is not a pure unit-root process but instead is the sum of a unit-root process and a cyclical (that is, business cycle) component. For instance, as Chart 2 shows, the two recessions seem to have depressed the growth rate of frequency, although the drop during the 1990/1991 recession appears to have been permanent (instead of cyclical).

Chart 2

Logarithmic Growth Rate of Frequency and First Difference in Rate of Unemployment, State-Level Data, Accident Years 1987-2004



Note: The Rate of Unemployment was measured in percent; for scaling purposes, the first difference was divided by 10 (in this exhibition only).

Table 1 displays the regression results for the local-level UC (severities) and STS (frequency) models. This table shows the final state variable only—the level μ_T. The t-statistic displayed alongside μ_T pertains to this final, time T variable only. Put differently, the t-statistic for μ_T does not afford statistical inference for μ_t in prior periods ($t = 1,...,T-1$). For medical severity, we can reject the null hypotheses of zero growth as at time T (the time of the last observation). Most interestingly, we can reject the null hypothesis that there is no business cycle influence on the growth rate of frequency. Specifically, an increase in the rate of unemployment by 1 percentage point (for instance, from 4 percent to 5 percent) depresses the (logarithmic) rate of growth of frequency by 1.76 percentage points. This finding supports the commonly held view that when the labor market softens, the least productive workers (which, frequently, are the last hired and thus least experienced) are the first to be laid off—such layoffs leaves the remaining pool of employed workers more experienced, on average. Note that, for the purpose of forecasting, the lack of statistical significance of "baseline" growth (as at time T) in indemnity severity and in frequency is irrelevant.

Table 1

Regression Results for Growth Rates of Indemnity and Medical Severities, and Frequency

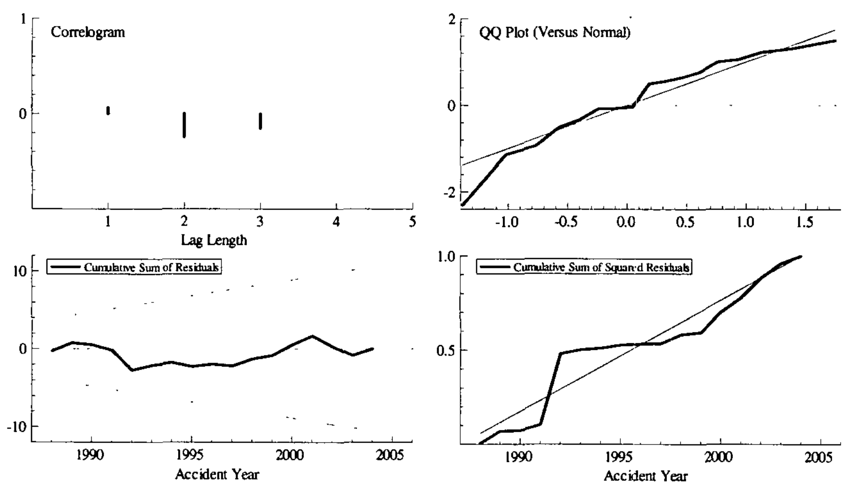
Variable	Coefficient	RMSE	t-statistic	Q-Ratio
Panel A: Indemnity Severity				
μ_T (Level)	0.020474	0.015641	1.309	0.0455
Log Likelihood	54.3136			
Panel B: Medical Severity				
μ_T (Level)	0.081500	0.014593	5.585	0.4060
Log Likelihood	59.8715			
Panel C: Frequency				
μ_T (Level)	0.0023129	0.0045367	0.50983	1.000
Unemployment	-0.017635	0.0074340	-2.3722	---
Log Likelihood	52.8787			

The Q-ratio in the rightmost column of Table 1 is the ratio of the ML-estimated variance of the innovation in level ($\hat{\sigma}_\eta^2$) to the ML-estimated variance of the measurement error ($\hat{\sigma}_\varepsilon^2$). For both severities and for frequency, this Q-ratio is positive. A positive Q-ratio indicates that the level μ_t is time-variant or, equivalently, that the rate of growth of the severity in question (for frequency, this holds net of the cyclical influence) is non-stationary, as hypothesized.

As mentioned, traditional measures of goodness of fit are of limited use for forecasting models. This impediment puts the emphasis on regression diagnostics. Chart 4 shows for the indemnity severity UC model four diagnostic plots for the measurement error. The left-hand side plot in the top panel of Chart 3 presents autocorrelations in the residuals at lag lengths 1 through 3. These autocorrelations appear to be small on this plot, thus lending support to the assumption that the measurement errors are independently distributed. On the right-hand side of the top panel, there is a QQ-plot. This QQ-plot indicates that there are no fat tails. Specifically, there is neither statistically significant skewness (which equals 0.8208) nor statistically significant excess kurtosis (which measures 0.1698). The bottom panel of Chart 2 displays the cumulative sum of residuals (left) and cumulative sum of squared residuals (right). The cumulative residuals signify no discernible positive serial correlation as these sums are well within the error cone. The cumulative sum of squared residuals indicates no material heteroskedasticity, thus suggesting that the assumption of a time-invariant variance of the measurement error is adequate. The corresponding residual diagnostics for the medical severity UC model and the frequency STS model are displayed in Charts 5 and 6. Here again, there is no statistically significant skewness (0.3505 and -0.4684, respectively) or excess kurtosis (-0.8213 and -0.5680, respectively).

Chart 3

Regression Diagnostics (Local Level UC Model) for the Log Growth Rate of Indemnity Severity



A means of testing the performance of the forecasting model is to generate forecasts for a holdout period and then compare these forecasts with the actual, known observations. We estimate the two severity models for the time period 1987 through 2001 (that is, periods $t = 1,...,T-3$), assigning the years 2002 through 2004 (periods $T-2$ through T) to the holdout window. Then, we generate multi-step logarithmic annual growth rate forecasts for this holdout window from the shortened ($t = 1,...,T-3$) time series. Multi-step forecasts, by definition, do not incorporate information that arrives during the holdout period; for instance, the forecast for T does not incorporate information that becomes available during periods $T-2$ or $T-1$. The concept of multi-step forecasting agrees with the actual forecasting problem at hand. However, there is one important difference between the holdout forecasting exercise and the actual forecasting situation. When we employ STS models in forecasting for the periods $T+1$ through $T+3$, we have to feed to the model for the first difference in the rate of unemployment historical observations or, if such observations have not yet become available, forecasts. In the holdout forecasting exercise, the historical observations for the state-level rate of unemployment in periods $T-2$ through T are available. Although using historical forecasts rather than historical observations in the holdout forecasting exercise would remedy this problem, the exercise would still be three observations short of the data series available in the actual forecasting situation.

Chart 4

Regression Diagnostics (Local Level UC Model) for the Log Growth Rate of Medical Severity

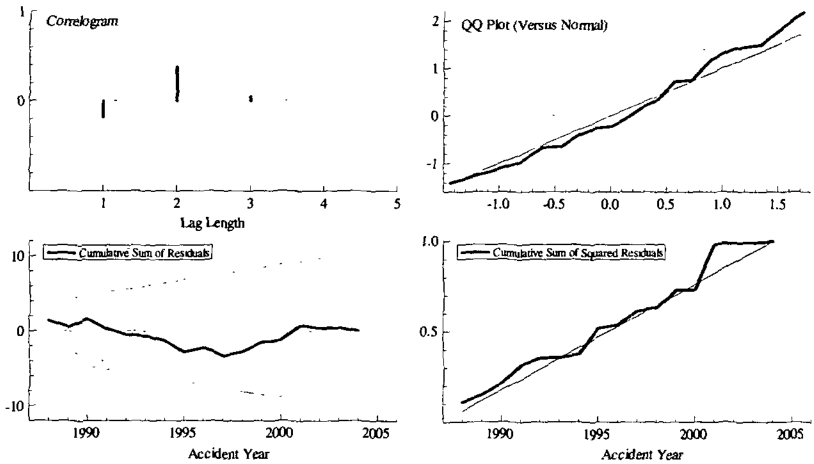
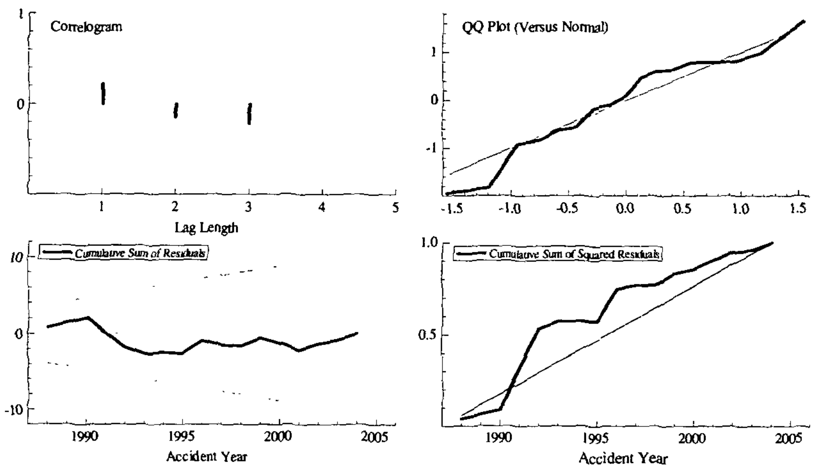


Chart 5

Regression Diagnostics (Local Level STS Model) for the Log Growth Rate of Frequency



In the actual forecasting situation, the accuracy of the forecasts for the frequency rate of growth is dependent on the accuracy of the forecasts for the rate of unemployment. (Note that our model does not account for the stochastic nature of the unemployment rate forecasts.) Most importantly, an STS model will forecast recession-related dips in the growth rate of frequency (as exhibited in the historical data of Chart 2) only if the pertinent forecasts for rate of unemployment describe such a recession. Unfortunately, economic recessions are next to impossible to forecast, because, if they were predictable, they would not occur as the Federal Reserve (or, possibly Congress, when it comes to fiscal policy) would act in a timely manner to prevent them.

Chart 6 and Chart 7 exhibit for the mentioned 3-year holdout window multi-step forecasts for the annual logarithmic growth rates of the indemnity and medical severities. Note that the displayed forecasts need to be multiplied by 100 to obtain percentage rates of growth. The confidence bounds around the forecasts range over 2 RMSE (root mean squared errors)—these confidence intervals are comparatively wide, which is due to the small number observations (14 by count). The forecasts of interest—those for the year $T+3$ (2004)—are quite accurate.

Chart 6

Holdout-Window Forecasts (Local Level UC Model) for the Growth Rate of Indemnity Severity

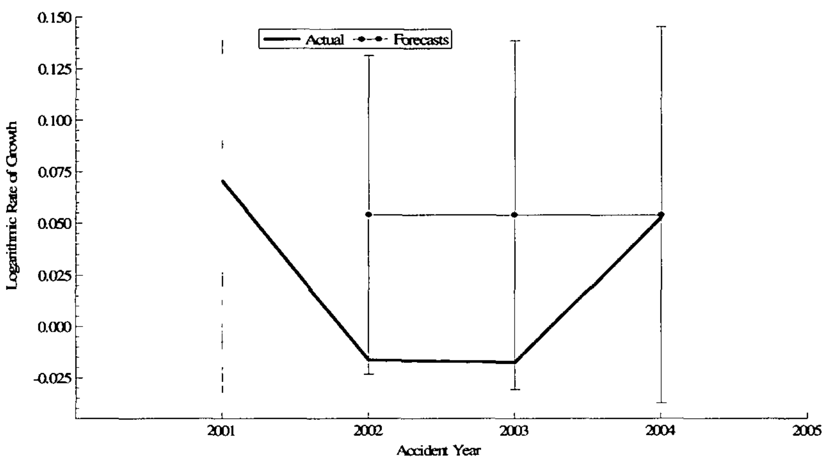
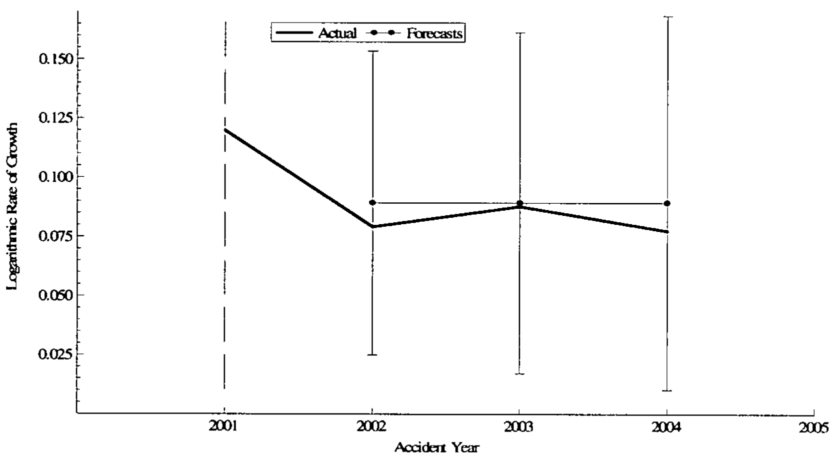


Chart 8 exhibits the hold-out window forecasts for the growth rate of frequency. The forecast of interest—the one for the year $T+3$ (2004)—clearly falls short of the observed value. Apparently, the regression coefficient that gauges influence of the first difference in the rate of unemployment underestimated the effect on the growth rate of frequency of the pronounced drop in the rate of unemployment rate during the economic recovery 2002-2004. It bears to mention that the shortening of the period of observation for this hold-out window exercise leaves the remaining period of observations (1987-2001) with only one economic recovery (the one following the 1990/91 recession); thus, it comes at now surprise that the regression coefficient in question is poorly estimated. Yet, it bears to mention that even in this shortened time period of only 15 observations, the STS model forecast for the year $T+3$ (2004) beats the benchmark forecast (the random walk), which equals -7.78 percent.

Chart 7

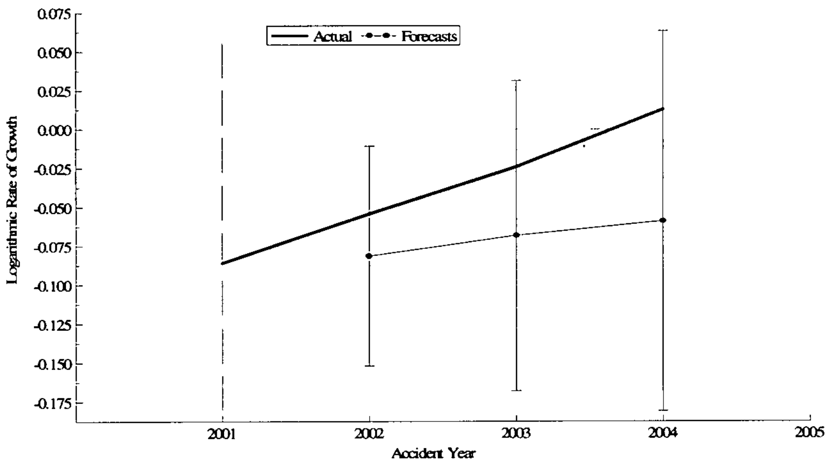
Holdout-Window Forecasts (Local Level UC Model) for the Growth Rate of Medical Severity



Charts 9 through 11 exhibit forecasts for the annual logarithmic rates of growth of the indemnity and medical severities and for frequency for the periods $T+1$ through $T+3$, based on the full ($t=1,...,T$) model. The top panels of these charts exhibit recent observations (at the dashed vertical bar and to the left of it) and forecasts with error bars ranging over 2 RMSE (solid vertical bars). Remember that the forecasts for frequency (unlike those for the severities) are linear combinations of the trend rate of growth (μ_T) and a business cycle component. The bottom panels of Charts 9 through 11 display the estimated (at the dashed vertical bar and to the left of it) and forecast values for the level μ_t. Note that these forecasts for the level—the annual trend rate of growth of the indemnity severity, medical severity, or frequency—are equal to the respective final state vector μ_T for all $T+f$, $f=1,...,\infty$. Again, the presented forecasts need to be multiplied by 100 to obtain percentage rates of growth.

Chart 8

Holdout-Window Forecasts (Local Level STS Model) for the Growth Rate of Frequency



As mentioned, the sought-after growth factor $1+g_{T,T+3}$ is the exponentiated sum of the 3 forecasts for the annual logarithmic rates of growth $\hat{g}^{\log}_{T+f}$ $(f=1,2,3)$.

Finally, it is of interest how the Kalman filter technique compares with less sophisticated approaches to forecasting logarithmic rates of growth from past realizations. For simplicity, we focus on medical severity as this variable is, unlike frequency, not (hypothesized to be) subject to the business cycle.

In general, if a time series is stationary (here, is the sum of a constant and a Gaussian error term), then the mean of the series renders unbiased forecasts for any future value of this series. Although any past value of the series renders such an unbiased forecast, the more past realizations are averaged over, the lower is the expected RMSE of this forecast.

Hence, if the log rate of growth of indemnity severity were stationary (which it is not, we hypothesize), then taking the mean over all available historical observations is desirable when forecasting this variable.

Chart 9

Forecasts (Local Level UC Model) for the Log Growth Rate of Indemnity Severity

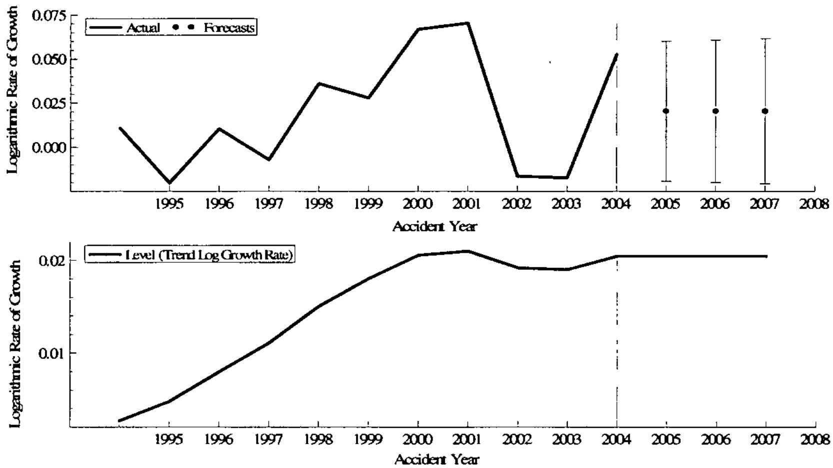


Chart 10

Forecasts (Local Level UC Model) for the Log Growth Rate of Medical Severity

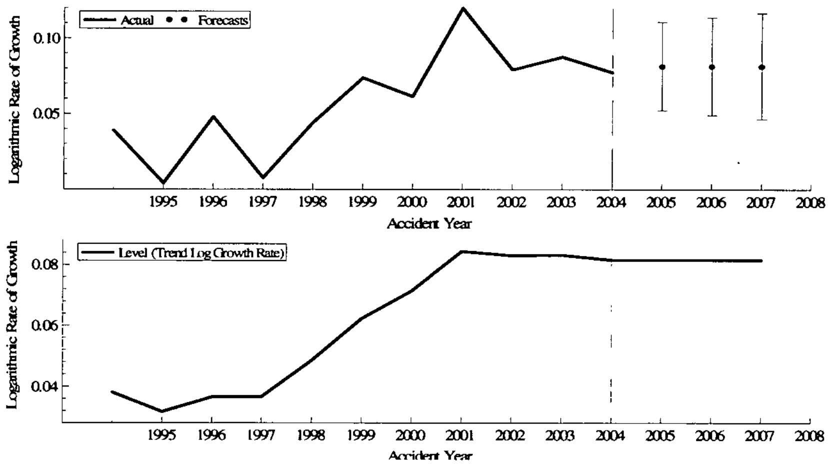
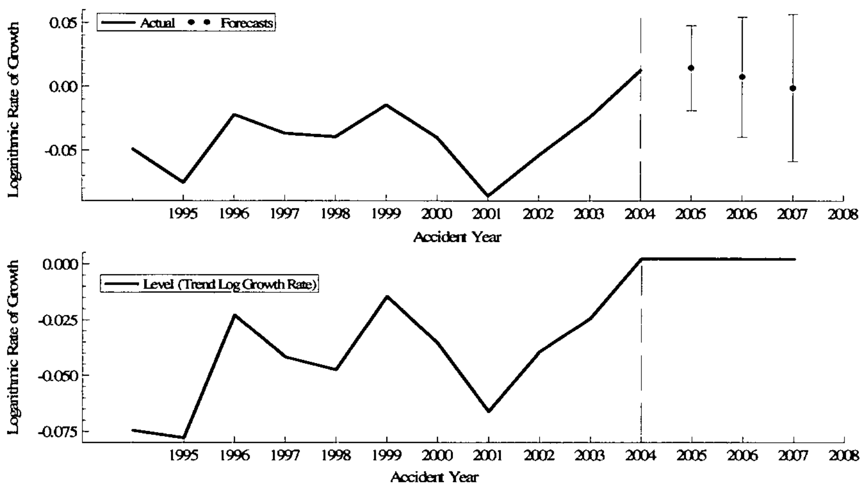


Chart 11

Forecasts (Local Level UC Model) for the Log Growth Rate of Indemnity Severity



If the growth rate of medical severity follows a random walk, as hypothesized above, then no averaging should be done. In general, if a time series follows a random walk, then the last (that is, the time T) realization serves as an unbiased forecast for any future value of this series, given the information available at time T. Hence, the last, time T observed log rate of growth for medial severity—the reading for the 2004 accident year is 7.74 percent—is an unbiased forecast for the log growth rates of periods $T+f$, $(f=1,2,3)$ (that is, accident years 2005 through 2007). Yet, as shown in Table 1, the final state vector and, hence, the Kalman-filtered forecast for the annual log rate of growth equals 8.15 percent. This 41 basis point annual difference is due to the measurement error in the observed data. Although both the last observed rate of growth (0.0774 in this realization) and the final state vector (0.0815 in this realization) reflect an unbiased forecast for any future period, these forecasts differ in precision. The Kalman-filtered forecast of 8.15 percent is likely to be much closer to the actual future outcome.

As a demonstration of the difference in typical outcome accuracy between the Kalman-filtered forecasts and forecasts that disregard a potential measurement error in the observed data, consider the holdout forecast for medical severity presented above. For the Kalman-filtered forecasts of the annual log rates of growth, the sum of the absolute forecast errors (for periods $T+1$, $T+2$, and $T+3$) equals 0.0387, and the RMSE amounts to 0.0090. When the last observed rates of growth are used, these gauges of forecast inaccuracy are as high as 0.1154 and 0.0234, respectively.

4. CONCLUSIONS

The experience of NCCI with Kalman filtered estimation of trend rates during the policy year 2006 rate filing season was encouraging. NCCI anticipates continued use of unobserved components models and structural time series models in future rate filings and the Kalman filter estimation technique. Current research at NCCI focuses on testing a Bayesian five-equation state-space forecasting model for severities, frequency, and the corresponding loss ratios—this multi-equation model, which accounts for add-up constraints and contemporaneous (cross-equation) covariances, is estimated using the Metropolis-Hastings algorithm.

Acknowledgment

We are grateful to Harry Shuford for comments. Thanks to Auntara De and Chris Doyle for research assistance. All errors are the authors'.

Equation Chapter 1 Section 1**5. REFERENCES**

[1] Bryson, Arthur E.; Ho, Yu-Chi, *Applied Optimal Control: Optimization, Estimation, and Control*, **1969**, Blaisdell.
[2] De Jong, Piet, "Smoothing and Interpolation with the State Space Model," *Journal of The American Statistical Association*, **1989**, Vol. 84, 1085-1088.
[3] Durbin, James; Koopman, Siem Jan, *Time Series Analysis by State Space Models*, **2001**, Oxford University Press.
[4] Green, William H., *Econometric Analysis*, 5th ed. **2002**, Prentice Hall.
[5] Hartwig, Robert P.; Kahley, William J.; Restrepo, Tanya E.; Retterath, Ronald C., "Workers Compensation and Economic Cycles: A Longitudinal Approach," *PCAS* **1997**, Vol. LXXIV, 660-700.
[6] Harvey, Andrew C., Forecasting, *Structural Time Series Models and the Kalman Filter*, **1991**, Cambridge University Press.
[7] Kalman, Rudolf E., "A New Approach to Linear Filtering and Prediction Problems," *Transactions ASME Journal of Basic Engineering*, **1969**, Vol. D82, 35-45.
[8] Kim, Chang-Jin; Nelson, Charles R., *State-Space Models With Regime Switching*, **1999**, MIT Press.
[9] Kohn, Robert; Ansley, Craig F., "A Fast Algorithm for Signal Extraction, Influence and Cross-validation in State Space Models," *Biometrika*, **1989**, Vol. 76, 65-79.
[10] Koopman, Siem Jan; Harvey, Andrew C.; Doornik, Jurgen A.; Shephard, Neil, *Stamp Structural Time Series Analyser Modeller And Predictor*, **1995**, Timberlake Consultants Ltd.
[11] Koopman, Siem Jan; Shephard, Neil; Doornik, Jurgen A., "Statistical Algorithms for Models in State Space Using SsfPack 2.2," *Econometrics Journal*, **1999**, Vol. 2, 113-166.
[12] Lommele, Jan A.; Sturgis, Robert W., "An Economic Model of Workmen's Compensation," *PCAS* **1974**, Vol. LXI, 170-189.
[13] McGuinness, John S., "Elements of Time-Series Analysis is Liability and Property Insurance Ratemaking," *PCAS* **1968**, Vol. LV, 202-254.
[14] Plato, *The Republic*, Book VII, **390 BCE**.
[15] Van Slyke, Oakley E., "Is Econometric Modeling Obsolete?" *Casualty Actuarial Society Discussion Paper Program* **1980**, 650-680.
[16] Venter, Gary G., "Classical Partial Credibility with Application to Trend," *PCAS* **1986**, Vol. LXXIII, 27-51.
[17] Zivot, Eric; Wang, Jeffrey; Koopman, Siem Jan, "State Space Modelling in Macroeconomics and Finance Using SsfPack in S+Finance," State Space and Unobserved Component Models, **2004**, Cambridge University Press.

Abbreviations and notations

ARIMA, auto-regressive integrated moving average
ARMA, auto-regressive moving average
ML, maximum likelihood
MLE, maximum likelihood estimation
NCCI, National Council on Compensation Insurance, Inc.
RMSE, root mean squared error
STS model, structural time series model
UC model, unobserved components model

Biographies of the Authors

Jonathan Evans, FCAS, MAAA is an Actuary at the National Council on Compensation Insurance, Inc.
Frank Schmid, Dr. habil., is a Director and Senior economist at the National Council on Compensation Insurance, Inc. He is also a 2006 Hicks-Tinbergen co-medalist for the best paper published in the Journal of the European Economic Association in the period corresponding to the years 2004 and 2005.