Bayesian Trend Selection

Frank Schmid, Chris Laws, and Matthew Montero

Abstract

Motivation. Selecting loss ratio trends is an integral part of NCCI aggregate ratemaking. The trend selection draws on an exponential trend (ET) regression model that is applied, alternatively, to the latest 5, 8, and 15 observations (dubbed 5-point, 8-point, and 15-point ET). Then, using actuarial judgment (which may account for a variety of influences), the three estimates are aggregated into a single forecast. This process of decision making under uncertainty can be formalized using Bayesian model selection.

Method. A Bayesian trend selection (BTS) model is introduced that averages across the three ET models. Using a double-exponential likelihood, this model minimizes the sum of absolute forecast errors for a set of (overlapping) holdout periods. The model selection is accomplished by means of a categorical distribution with a Dirichlet prior. The model is estimated by way of Markov chain Monte Carlo simulation (MCMC).

Results. The BTS is validated on data from past ratemaking seasons. Further, the robustness of the model is examined for past ratemaking data and a long series of injury (and illness) incidence rates for the manufacturing industry. In both cases, the performance of the BTS is compared to the 5-point, 8-point, and 15-point ET, using the random walk as a benchmark. Finally, for the purpose of illustration, the BTS is implemented for an unidentified state.

Availability. The model was implemented in R (cran.r-project.org/), using the sampling platform JAGS (Just Another Gibbs Sampler, www-ice.iarc.fr/~martyn/software/jags/). JAGS was linked to R via the R package rjags (cran.r-project.org/web/packages/rjags/index.html).

Keywords. Model Selection, Model Averaging, Model Robustness, Aggregate Ratemaking, Trend Estimation, Workers Compensation

1. INTRODUCTION

Selecting loss ratio trends is an integral part of NCCI aggregate ratemaking. The trend selection draws on an exponential trend (ET) regression model that is typically applied to the latest 5, 8, and 15 observations; these three alternative regressions are dubbed the 5-point, 8-point, and 15-point ET. Then, using actuarial judgment (which may account for a variety of influences), the three estimates are aggregated into a single forecast. This process of decision making under uncertainty can be formalized using Bayesian model selection.

In what follows, a Bayesian trend selection (BTS) model is introduced that averages across the three ET models. Using a double-exponential likelihood, this model minimizes the sum of absolute forecast errors for a set of (overlapping) holdout periods. The model selection is accomplished by means of a categorical distribution with a Dirichlet prior. The model is estimated by way of Markov chain Monte Carlo simulation (MCMC).

The BTS is validated on data from past ratemaking seasons. Further, the robustness of the model is examined for past ratemaking data and a long series of injury (and illness) incidence rates for the manufacturing industry. In both cases, the performance of the BTS is compared to the 5-point, 8-point, and 15-point ET, using the random walk as a benchmark. Finally, for the purpose of illustration, the BTS is implemented for an unidentified state.

The BTS, as any model, is limited in the scope of information that it is capable of processing—at the same time, this statistical model is not subject to potential biases of human decision making. On one hand, human judgment is capable of considering influences beyond the scope of statistical models, such as changing economic conditions. On the other hand, research pioneered by Amos Tversky and Daniel Kahneman shows that human decision making is subject to systematic errors. Although decision making under uncertainty is liable to random errors (if only due to measurement errors in the data), systematic errors lead to predictable biases. Most interestingly, Tversky and Kahneman [10] argue that the prevalence of biases is "not restricted to the layman. Experienced researchers are also prone to the same biases—when they think intuitively."

Among the biases identified by Tversky and Kahneman [10] is one related to the availability of events, that is, the ease with which occurrences can be retrieved from the memory bank. The more available an event is, the more this event weighs on the decision, all else being equal. Among the factors shaping the availability of an event is its salience. Events that occur more recently or have greater degrees of immediacy (for instance, seeing a house burn, as opposed to reading about the fire in the newspaper) loom larger on the human mind. The salience bias poses a risk of overreacting to recent events and extreme outcomes. For an extensive discussion of the availability bias, see Taylor [9].

1.1 Research Context

Past work on trend modeling in the context of ratemaking in workers compensation includes Brooks [2], Evans and Schmid [4], and Schmid [7]. All of these approaches face a major challenge in the shortness and comparatively high volatility of the available time series.

Brooks [2] studied the determinants of indemnity frequency for California using ordinary least squares regression. Indemnity frequency is defined as the ratio of indemnity-related claim count to payroll at the 1987 wage level. The explanatory variables comprise measures of indemnity benefit and medical benefit levels, macroeconomic variables, and the ratio of cumulative injury claims to

total indemnity claims; the latter variable accounts for more than half the explanatory power of the presented models.

In Brooks [2], all variables are transformed into first differences (on the raw scale; on the scale of the natural logarithm, such first differences would represent continuously compounded rates of growth). The author runs 84 indemnity frequency regressions (the maximum possible number, given the set of available explanatory variables), orders them by the adjusted R-squared, and then publishes the seven highest ranking models.

Brooks [2] aims at a high R-squared (adjusted for the degrees of freedom), by having "accounted for as much variation as possible." Brooks [2] reads a high explanatory power of the model as an indication that the estimates "are not distorted due to a misspecified model with a large portion of unaccounted-for variance." Yet, there is little relation between a small forecast error and a high R-squared. For instance, in a game of (two six-faced) dice, seven is the best forecast for any toss, because this is the forecast that minimizes the expected forecast error; at the same time, regressing 49 sevens on the 49 possible outcomes of this game of dice generates an R-squared of zero. See also Armstrong [1], who argues against the use of the R-squared (and, more generally, measures of in-sample fit) in selecting the most accurate time-series model.

The approach of considering a multitude of models, as pursued by Brooks [2], creates a condition of multiple comparisons. Considering a set of statistical hypotheses simultaneously increases the probability of false positives in the context of traditional significance tests, and thus invalidates conventional measures of statistical significance. For instance, when applying a type 1 error probability of 10 percent in an *F*-test (that is, in an analysis of variance, which tests the statistical significance of the model), 10 percent of a random set of models are statistically significant by chance, on average. Specifically, among the 84 models that Brooks [2] ran, in expected value terms, eight will turn out statistically significant by chance.

Evans and Schmid [4] discuss the use of the Kalman filter for frequency and severity forecasting in NCCI aggregate ratemaking. Although, in general, the Kalman filter is a powerful tool for breaking down time-series data into signal and noise, the time series employed in NCCI ratemaking are too short to estimate reliably the variances of innovation and white noise. Specifically, in short time series, the Kalman filter may overestimate the innovation variance, thus favoring the random walk (over alternative data-generating processes).

Schmid [7] tries to overcome the problem associated with the Kalman filter in short time series by using a random-walk smoother. The innovation variance of this smoother is calibrated such that the root mean squared forecast error for the time horizon of three years is minimized, on average. Although this model performs as expected, it is difficult to communicate to the decision maker.

The advantage of the approach by Brooks [2] is its use of covariates, which, in a single-equation model, may improve the forecasts considerably. Forecasts for macroeconomic variables that may serve as covariates may be readily available from professional forecasting firms. Conversely, the model developed by Brooks [2] requires forecasts for the ratio of cumulative injury claims to total indemnity claims, which is the covariate that wields the most explanatory power (Brooks [2]). This ratio may be just as difficult to forecast as the variable of interest. From this perspective, the use of covariates can pose an obstacle when using single-equation regression models for the purpose of forecasting.

1.2 Objective

The objective is to formalize the actuarial decision making process of trend selection, rather than developing a new trend model. Statistically, this decision making process under uncertainty is a problem of model selection. The models that compete in this selection process are the 5-point, 8-point, and 15-point ET. The resulting forecast is a weighted average of the three ET estimates, where the weights are the posterior probabilities generated by the model selection process.

1.3 Outline

What follows in Section 2 is a description of the forecasting task and the data. Section 3 offers the theoretical framework of Bayesian model selection. Section 4 presents the validation of the model on past ratemaking data. Section 5 examines the robustness of the model; in addition to ratemaking data, the model is applied to a long time series of manufacturing injury (and illness) incidence rates. Section 6 applies the model to an unidentified state. Section 7 concludes.

2. FORECASTING TASK AND DATA

The objective is to forecast, in the context of aggregate ratemaking, compound annual growth rates for the indemnity and medical loss ratios or, alternatively, compound annual growth rates for frequency and the indemnity and medical severities. Severity is defined as the ratio of (on-leveled, developed-to-ultimate, and wage-adjusted) losses to the number of lost-time claims (developed to ultimate). Frequency is defined as the ratio of the lost-time claim count (developed to ultimate) to

on-leveled and wage-adjusted premium. The loss ratio is the product of frequency and the respective severity.

The forecasts apply to the time period between the midpoint of the respective policy (accident) year included in the experience period and the midpoint of the proposed rate effective period—this time interval is called the trend period. For the latest policy (accident) year that is included in the experience period, the length of the trend period usually amounts to slightly more than three years. The rate effective period is the time period during which the rates are expected to be effective, based on the rate filing. In NCCI rate filings, the experience period comprises at least two policy (accident) years.

All data are state-level observations and are on an annual basis. From the perspective of the model, there is no difference between processing policy year and accident year data. (In NCCI ratemaking, nearly all states operate on policy year information.)

The ET and, hence the BTS, do not make use of covariates or an autoregressive process. As a result, the forecast compound annual growth rate (CAGR) can be scaled up to the trend period simply by means of compounding.

In NCCI ratemaking, where the experience period comprises more than one policy (or accident) year, the ET models make use of policy (or accident) year observations through the end of the experience period. Then, different scale factors are applied for the individual policy (or accident) years in the experience period when compounding the CAGR.

3. STATISTICAL MODELING

The BTS is a tool for decision making under uncertainty. Statistically, this concept of decision making is implemented by means of model selection in a Bayesian framework. The objective of the approach is to arrive at a forecast by means of weighted averaging across a set of candidate models. This set of models arises from the ET applied to a set of time intervals, all of which end with the most recent observation but differ by the degree to which they extend into the past. The weights of the average originate in the posterior probabilities by which each of the approaches is deemed to be the "true" model.

3.1 Exponential Trend (ET)

The ET model regresses the (natural) logarithm of (for instance) frequency on a linear trend. The forecast CAGR is then backed out of the regression coefficient that captures this linear time trend.

The semi-logarithmic regression equation reads

$$\ln(Y_t) = \alpha + \beta \cdot t + \varepsilon_t \ , \ t = 1,...,T \ , \tag{1}$$

where $\ln(Y_t)$ is the dependent variable (e.g., the natural logarithm of frequency), α is the intercept, β is the parameter that captures the time trend, ε_t is an error term, and T equals the number of observations of the time series.

The forecast CAGR equals

$$e^{\hat{\beta}} - 1, \tag{2}$$

where $\hat{\beta}$ is the ordinary least squares estimate of the parameter that reflects the influence of time.

Note that the least squares approach makes no assumption about the distribution of the conditional mean of the dependent variable. In particular, the conditional mean of y need not be normally distributed for the ET to deliver meaningful estimates.

Due to the loss ratio being equal to the product of frequency and the respective severity, the following relation holds for the CAGR forecasts:

$$e^{\hat{\beta}_{\text{loss ratio}}} - 1 = \left(e^{\hat{\beta}_{\text{frequency}}} \right) \cdot \left(e^{\hat{\beta}_{\text{severity}}} \right) - 1 \ . \tag{3}$$

In NCCI ratemaking, the ET is typically applied to the latest 5, 8, and 15 data points. The forecast growth rate is judgmentally determined and may be interpreted as a weighted average of the resultant three CAGR. By applying judgment, it is possible to factor in influences that lie outside the model, such as additional trend estimates or changes in economic conditions.

3.2 Bayesian Trend Selection (BTS)

In what follows, the nature of the BTS approach is illustrated for the frequency trend selection; the loss ratio trend selections are implemented accordingly and independently. The CAGR for the severity rates of growth are then backed out of the CAGR of frequency and the respective loss ratio. Loss ratio modeling is given preference over the modeling of severity, since loss ratio rates of growth are generally smoother than severity (and frequency) rates of growth. This is because the numerator of frequency equals the denominator of severity and, as a result, some of the variation of the frequency and the severity series cancels out at the level of the loss ratio.

In a first step, for a given state, three values for the CAGR of frequency are estimated using the 5-point, 8-point, and 15-point ET. A holdout period of three years applies to each of the three ET specifications, which means that the time windows of the ET models shift into the past by three

years. This estimation process is performed for the S most recent ratemaking data sets (inclusive of the ongoing ratemaking season). Currently, S equals three but may be allowed to increase with future ratemaking seasons. By employing such a holdout period, the CAGR forecasts can be compared to the realized values three years hence. The three-year holdout period is motivated by the length of the trend period in aggregate ratemaking, which typically amounts to slightly more than three years, as mentioned. (Note that the S holdout periods are overlapping, since they are three-year time windows that move in increments of one year.) The realized CAGR three years hence that is compared to the ET forecast defined in Equation (2) is calculated as $\sqrt[3]{Y_T / Y_{T-3}} - 1$, where Y is the observed value of frequency.

In the BTS framework, the forecasts and the realized values three years hence are related to each other by means of a double exponential distribution. In this model, there are S observations, since every frequency data set generates one data point. The three ET forecasts compete for being considered as the location parameter of the double-exponential distribution. This selection process is accomplished by means of a categorical distribution with a Dirichlet prior.

In relating forecasts to realized values, the double exponential distribution minimizes the sum of absolute forecast errors, as opposed to the sum of squared forecast errors. In the context of the BTS, minimizing the sum of squared forecast errors would be equivalent to minimizing the root mean squared forecast error, which Armstrong [1] has shown to be "unreliable, especially if the data might contain mistakes or outliers." Instead, by minimizing the sum of absolute forecast errors, the BTS is robust to outliers.

The BTS model reads

$$y_s \sim \text{Dexp}(\mu_s, \tau),\ s = 1,...,S, \tag{4}$$

$$\mu_s = \vec{y}_s^{\,f}\,[\lambda],\ s = 1,...,S, \tag{5}$$

$$\lambda \sim \text{Cat}(\vec{p}), \tag{6}$$

$$\vec{p} \sim \text{Dirch}(\vec{\alpha}), \tag{7}$$

$$\vec{\alpha} = (1,1,1), \tag{8}$$

$$\tau \sim \text{Ga}(0.001, 0.001), \tag{9}$$

$$\sigma = \sqrt{2}/\tau, \tag{10}$$

where the suffix $s = 1,...,S$ indicates the respective ratemaking data set. The most recent data set pertains to the ongoing ratemaking season, whereas the remaining $S-1$ data sets originate from ratemaking seasons of the immediate past.

Equation (4) presents the double exponential distribution, which defines the likelihood of the Bayesian model; as mentioned, a double exponential likelihood minimizes the sum of absolute errors, which are the S forecast errors from as many holdout periods.

The model selection mechanism comprises Equations (5) through (8). In Equation (5), the conditional mean of the double exponential distribution is chosen from the vector of the three competing CAGR forecasts $\vec{y}^{\,f}$ that arise from as many ET models. The parameter λ, which takes on integer values between (and inclusive of) unity and three, selects the element of the vector $\vec{y}^{\,f}$ that is to serve as the location parameter of the double exponential distribution. This parameter λ is generated by a categorical distribution (Equation [6]) with a Dirichlet prior (Equation [7]). The parameters of the Dirichlet distribution (Equation [8]) are identical, thus affording each of the three ET models equal prior probability of being the "true" model.

Equation (9) depicts the gamma prior for the scale parameter of the double exponential distribution. Equation (10) denotes the standard deviation of the double exponential errors, as an informational item.

The model is estimated by means of MCMC. Every iteration of the Markov chain generates a draw from the categorical distribution. The relative frequency with which a model is selected equals the posterior probability of this model representing the "true" data-generating process. The CAGR for frequency (and, similarly, for the loss ratios) is then computed as a weighted average of the CAGR values of the three ET models estimated without a holdout period, using the latest available observations. The weights of this average are the relative frequencies of the values assumed by the parameter λ. The CAGR values for the severities are backed out of the CAGR values for frequency and the loss ratios according to Equation (3).

Finally, in a sensitivity analysis, the BTS is re-estimated S times, each time leaving out one of the S data sets. This leave-out-one cross-validation (CV) generates a range of forecasts that do not necessarily include the original forecast (that was obtained when including all S data sets). Further, this range of forecasts can be comparatively narrow, because only the posterior probabilities (of the ET models) are cross-validated, not the ET estimates themselves. A CV range that lies below the BTS forecast indicates that the risk is skewed toward the downside, thus implying that the decision

maker should be cautious about selecting a value greater than the BTS forecast. Similarly, a CV range that lies entirely above the BTS forecast indicates a risk of the forecast being too low.

The JAGS code of the BTS outlined in Equation (4) through Equation (10) is documented in Appendix 8.1.

4. MODEL VALIDATION

The BTS is validated by means of comparing its forecast performance to the 5-point, 8-point, and 15-point ET. The forecasts of all four models are benchmarked to the random walk. If a time series follows a random walk, then the best forecast for any future time period is the latest observed value.

Implementing the BTS requires S data sets of a contiguous time period—the ongoing and past $S-1$ ratemaking seasons. As discussed, S is set to three but may be increased as the number of available data sets grows with future ratemaking seasons.

Validating the BTS necessitates a total of $S+3$ data sets. For instance, applying the BTS during the 2012 ratemaking season (where the latest available policy year is 2010) necessitates data from the 2011 and 2010 ratemaking seasons as well. Model validation requires, in addition, data sets from the ratemaking seasons 2007 through 2009.

In a first step, the BTS is applied retrospectively to the 2008 ratemaking season, thus making use of the 2008, 2007, and 2006 ratemaking data sets. Then, in a second step, the three-year CAGR forecasts associated with the three ET models and the BTS are compared to the CAGR that were observed for the time period 2009 through 2011. For this purpose, the sums of absolute forecast errors are computed across the set of analyzed states and normalized by the forecast error associated with the random walk. (For every state, there is one three-year forecast for each model and data series; for instance, there is one BTS forecast for the three-year CAGR of frequency.)

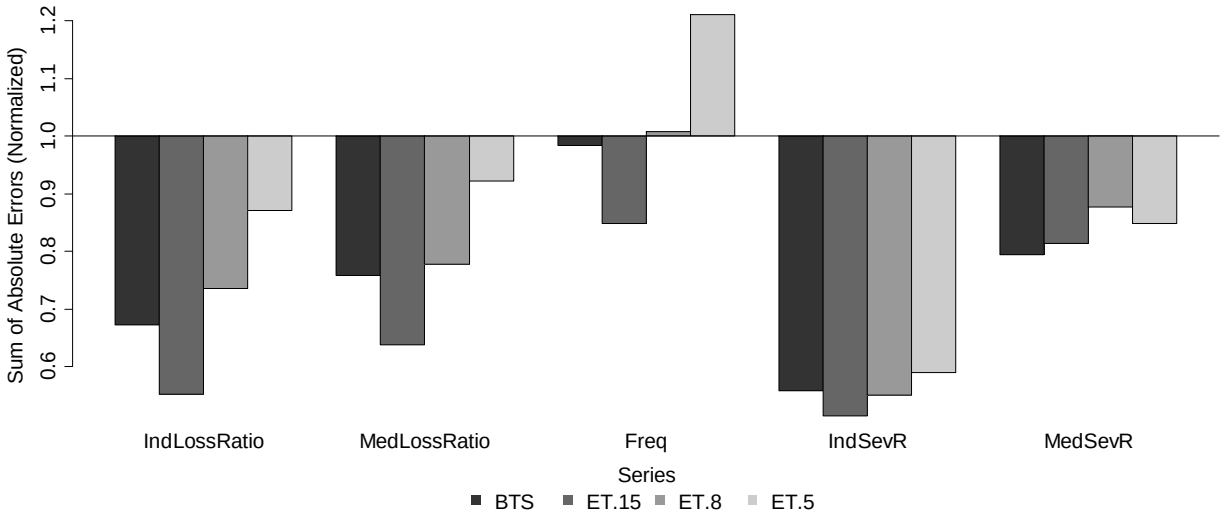
The model validation makes use of 29 states. States with an insufficient number of observations or insufficient number of past data sets could not be considered. Such data limitations may arise in states that switched from accident year to policy year information in the recent past or in states that have only recently been added to the data collection process.

Chart 1 depicts the absolute forecast errors, summed up across the 29 states and normalized by the forecast error of the random walk. Values greater than unity imply that the respective model underperforms the random walk, thus adding no information that is not yet comprised in the latest

observed data point. As judged by the loss ratio forecasts, the BTS is the second-best model, behind the 15-point ET. As an informational item, the BTS comes in first for medical severity, second for frequency, and third for indemnity severity, each time outperforming the random walk.

The performance of the ET models relative to the random walk is most interesting. On one hand, it appears that the frequency rate of growth is highly nonstationary, given the performance of the random walk compared to the BTS and the ET models. On the other hand, the only model that beats the random walk is the 15-point ET, which is the model that is most appropriate when the frequency growth rate is stationary.

Chart 1: Sum of Absolute Forecast Errors, Validated on the 2011 Ratemaking Season



5. ROBUSTNESS

The concept of model robustness has been pioneered by Lars Hansen and Thomas Sargent. According to this principle, a decision rule should be designed such that it "works well enough (is robust) despite possible misspecification of its model" (Ellison and Sargent [3]). In particular, the performance of the model should be adequate even in a worst-case scenario. For an extensive treatment of the concept of robustness, see Hansen and Sargent [5].

In the context of the BTS, robustness has two dimensions. First, the performance of the BTS in forecasting the CAGR of the loss ratios should be adequate for any state. Specifically, for the loss

ratios, the largest absolute forecast error of the BTS should be smaller than the largest absolute forecast error within the set of four alternative models, that is, the three ET models and the random walk. Chart 2 shows that the BTS satisfies this condition although, for the indemnity loss ratio, the advantage of the BTS over the worst-performing ET (5-point) and the random walk is slender. Regarding the ratio that defines the normalized largest absolute forecast error in Chart 2, numerator (model) and denominator (random walk) may not refer to the same state.

Second, the performance of the BTS should be adequate in an environment where the nature of the time series changes. As shown by Schmid [8], the data-generating process of the injury (and illness) incidence rate in the manufacturing industry has changed over the course of the 20th century from one of high variance and small serial correlation in the error term to one with low variance and high persistence in the error. The structural break in the degree of serial correlation occurred in the early 1960s and coincided with the end of a two-decade decline in the variance. In part, the change in the nature of the manufacturing incidence rate may be related to changes in the data collection process, although it is worthy of note that OSHA (Occupational Safety and Health Administration) did not become effective until 1971.

Chart 2: Maximum Absolute Forecast Error, Validated on the 2011 Ratemaking Season

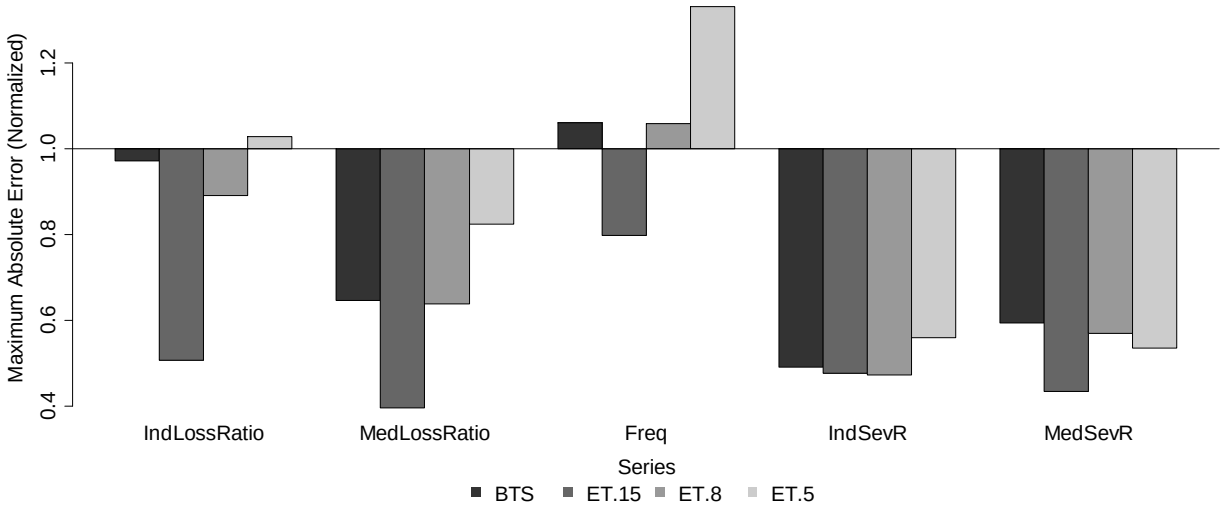


Chart 3 depicts the incidence rate of workplace injuries (and illnesses) for the manufacturing industry; the series runs from 1926 through 2010. The incidence rate is on a logarithmic scale, which implies that a straight line signifies a constant rate of growth. The gray vertical bars indicate

economic recessions, as defined by the NBER (National Bureau of Economic Research). With OSHA becoming operational in 1971, the incidence rate started including work-related illnesses. For methodological details on the incidence rate and for data sources, see Appendix 8.2.

Chart 3: Manufacturing Injury (and Illness) Incidence Rate, 1926–2010, per 100 FTE Employees

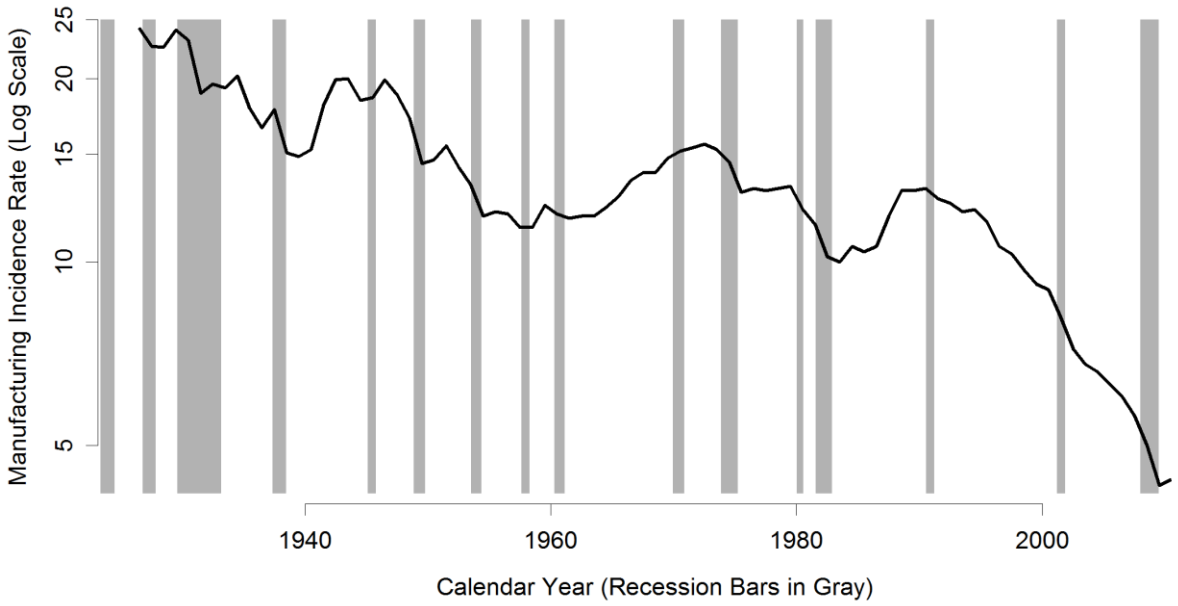


Chart 4 presents the logarithmic rates of growth (that is, first differences in natural logarithms) of the manufacturing injuries (and illnesses) incidence rate; again, the gray bars indicate recessions. The decline in variance and the increase in persistence of variations around zero (as a proxy for the negative, possibly time-varying mean) are apparent. Starting in the early 1960s, frequency increases are more likely to be followed by further increases than they were in the first half of the 20th century; likewise, frequency decreases show more persistence as well. Clearly, the nature of the time series changed over time.

In the context of the manufacturing incidence rate, a defining characteristic of a robust decision rule is its adequate forecast performance in both the early high-variance low-persistence regime and the later low-variance high-persistence regime. In what follows, the three ET models and the BTS are applied to the full series (1926–2010); the early regime, inclusive of the transition period (1926–1964); and the second regime.

Bayesian Trend Selection

Similar to studying model robustness for ratemaking data, the 5-point, 8-point, and 15-point ET models are applied with a holdout period of three years, using time windows that roll forward at increments of one year. The three forecast errors that are associated with the three neighboring time windows are the basis for the model selection of the BTS. The BTS forecast is then calculated as a weighted average of the 5-point, 8-point, and 15-point ET forecasts, where the ET models have been run with data leading up to the start of the forecasting horizon (and, hence, without the three-year holdout period). The resulting BTS forecast error is compared to the forecast errors of the 5-point, 8-point, and 15-point ET that enter the weighted average of the BTS forecast.

Chart 4: Manufacturing Injury (and Illness) Incidence Rate, Log Growth Rate, 1927–2010

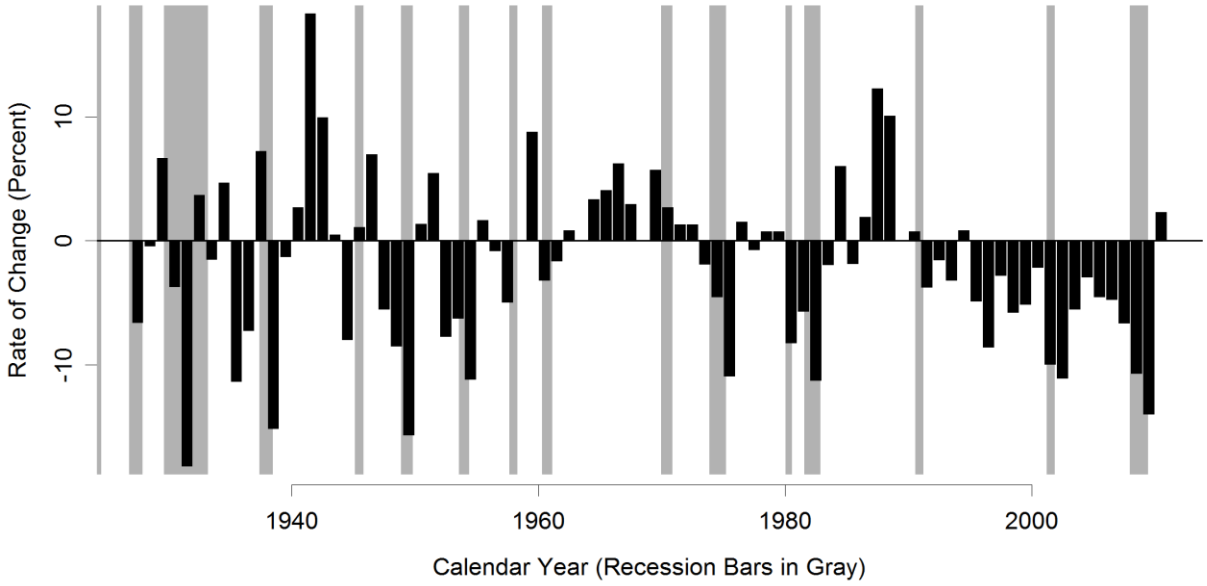


Chart 5 depicts for the full manufacturing incidence rate series the sum of absolute errors of the three ET models and the BTS, normalized by the sum of absolute forecast errors of the random walk. The BTS and the 5-point ET perform about equally well, outperforming the random walk and the 15-point and 8-point ET models, although the performance differences to the 8-point ET is slim.

Chart 5: Manufacturing Incidence Rate, Sum of Absolute Errors, Normalized, 1926–2010

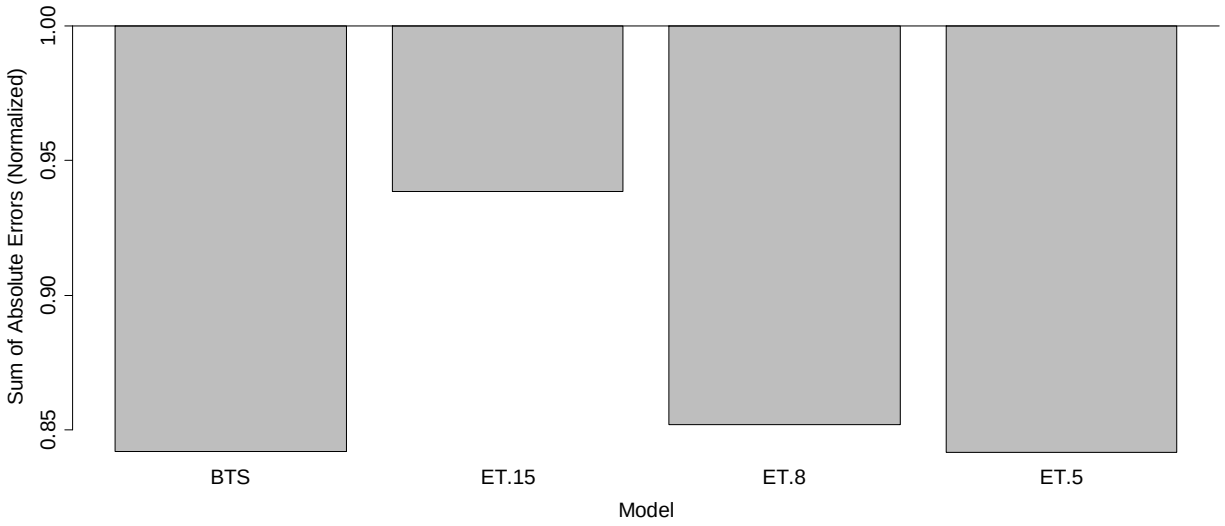
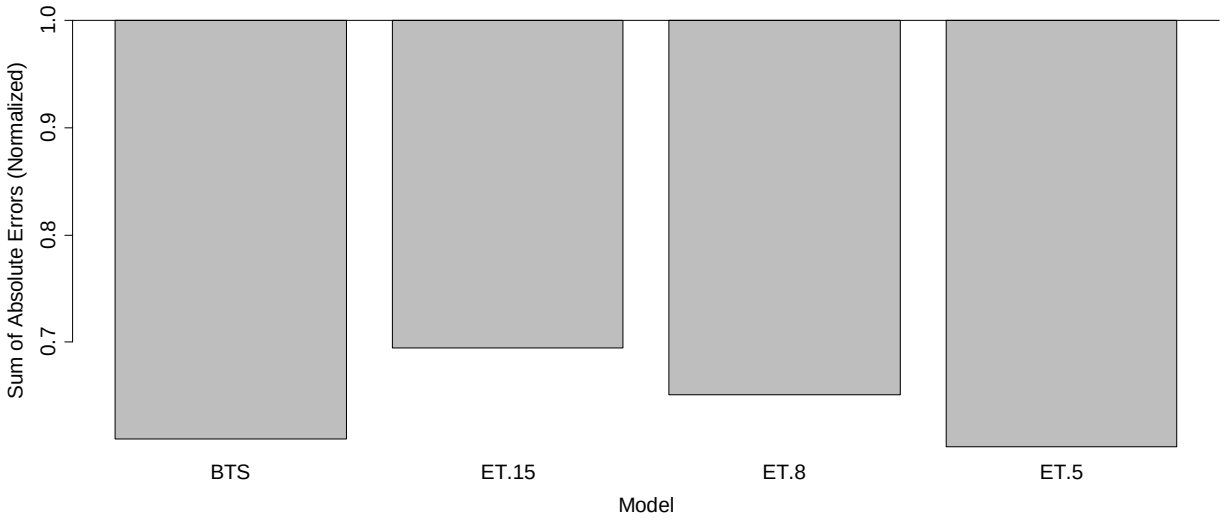


Chart 6 presents the analysis of the forecast error for the first, high-variance low-persistence regime and the subsequent transition period (1926–1964). Here again, the BTS and the 5-point ET perform about equally well, beating the random walk and the 15-point and 8-point ET models.

Chart 6: Manufacturing Incidence Rate, Sum of Absolute Errors, Normalized, 1926–1964

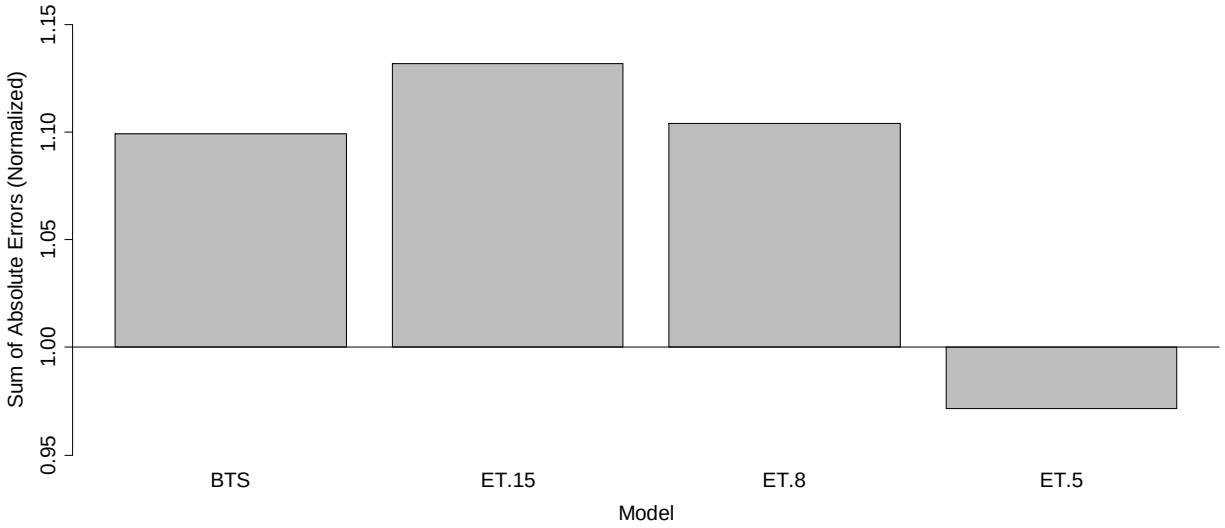


Finally, Chart 7 offers the analysis for the second regime (1965–2010). In this low-variance high-persistence setting, only the 5-point ET outperforms the random walk. The BTS comes in second, although the performance differences to the 8-point and 15-point ET models are slender.

Whereas, for the ratemaking data, the 15-point ET emerged as the highest performing model, for the manufacturing incidence rate, it is the 5-point ET. At the same time, the BTS outperforms the 15-point ET for the manufacturing series and beats the 5-point ET on the ratemaking data. The BTS performs well in both environments. Although the BTS never emerges as the winner, it proves to be the most robust model.

The BTS delivers a robust decision rule due to its ability to draw on any of the ET models; the BTS makes its model selection based on recent forecast performance. As the nature of a time series changes, the BTS adapts and leans toward the most suitable model. At the same time, the BTS is conservative in its model selection in that it never unequivocally adopts the highest performing model (which explains why it underperforms the 15-point ET on the loss ratios and the 5-point ET on the manufacturing incidence rate series).

Chart 7: Manufacturing Incidence Rate, Sum of Absolute Errors, Normalized, 1965–2010



6. APPLICATION

The BTS is applied to an unidentified state. The data originates from the 2011 NCCI ratemaking season.

There are two types of charts. The first type of chart displays the observed growth rates, the 5-point, 8-point, and 15-point ET estimates, and the BTS forecast. To avoid clutter, the CAGR values associated with the three ET models are indicated by boxes (instead of lines), where the horizontal distances between the boxes indicate the time intervals associated with 4, 7, and 14 rates of growth. Further, there are the CV forecast ranges, which may be close to vanishing in some of the charts.

The second type of chart displays the posterior probabilities related to the three ET models. As mentioned, these posterior probabilities are the relative frequencies with which any of the three ET models was selected during the BTS estimation process. This second type of chart is available only for frequency and the loss ratios—this is because the CAGR forecasts for the severities are backed out of the CAGR forecasts for frequency and the respective loss ratio.

Chart 8 depicts the analysis of the indemnity loss ratio, which displays a clear upward drift. The 5-point ET estimate exceeds the 8-point ET estimate, which, in turn, exceeds the 15-point ET estimate. As a consequence of this apparent nonstationarity (that is, time-variation in the mean rate of growth), the BTS affords high probabilities to ET models that do not reach deep into the past (see Chart 9).

Chart 10 presents the results for the medical loss ratio. The growth rate of this time series is comparatively stable, as indicated by the close proximity of the 5-point, 8-point and 15-point ET estimates to each other. As a result, the BTS leans toward stationarity (that is, the presumption of a time-invariant mean rate of growth), thus preferring estimates based on long time series to estimates that rely on only the most recent data points (see Chart 11). Clearly, if a time series is stationary, the CAGR should be calculated from a high number of data points, since this increases the accuracy of the estimate.

Chart 12 displays the frequency analysis. The rate of frequency growth appears to drift up, as indicated by the observation that the CAGR forecasts of the 8-point and 5-point ET models exceed the estimate of the 15-point ET. The BTS forecast is close to the 5-point and 8-point ET estimates, as implied by the high posterior probabilities afforded to the two models (see Chart 13).

Chart 8: Indemnity Loss Ratio, Growth Rates, 1996–2009

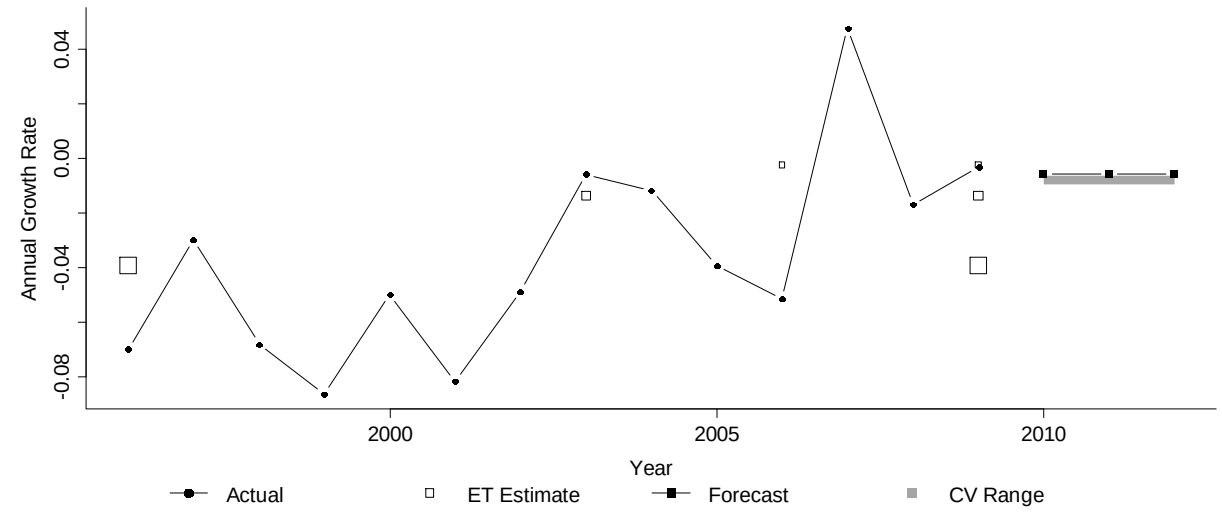


Chart 9: Indemnity Loss Ratio, Posterior ET Probabilities

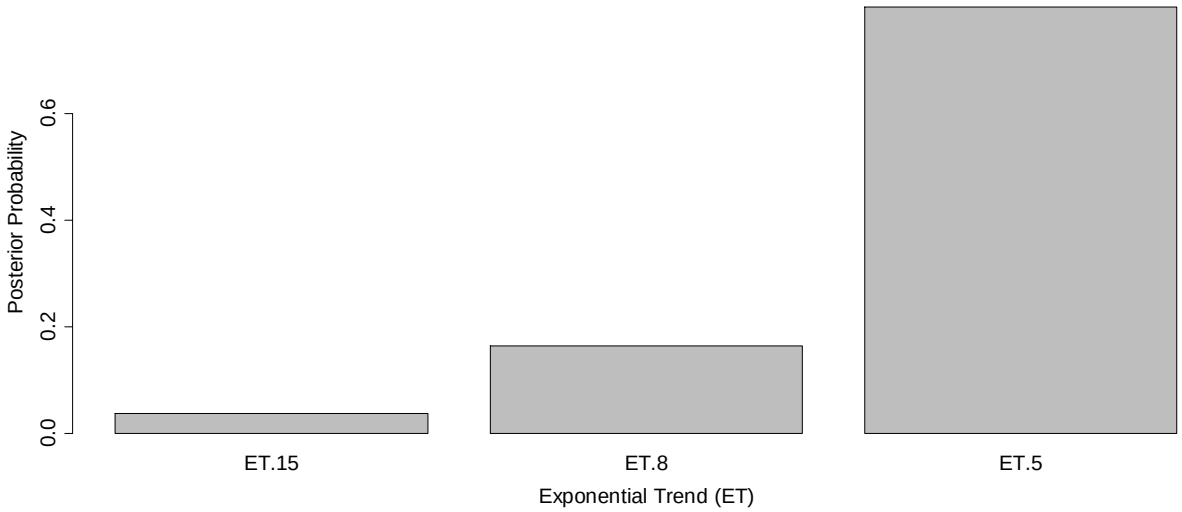


Chart 10: Medical Loss Ratio, Growth Rates, 1996–2009

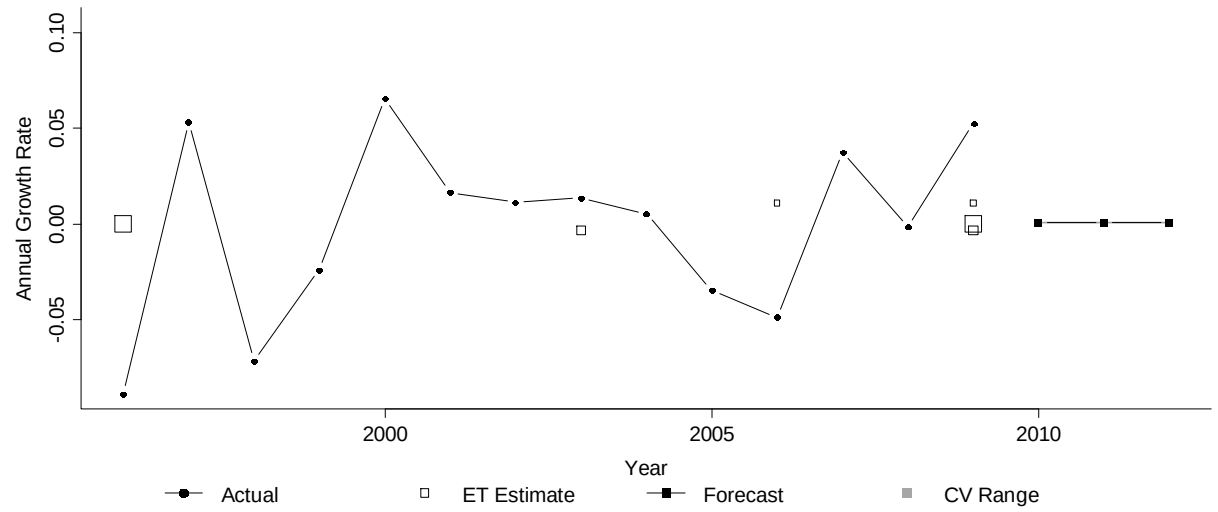


Chart 11: Medical Loss Ratio, Posterior ET Probabilities

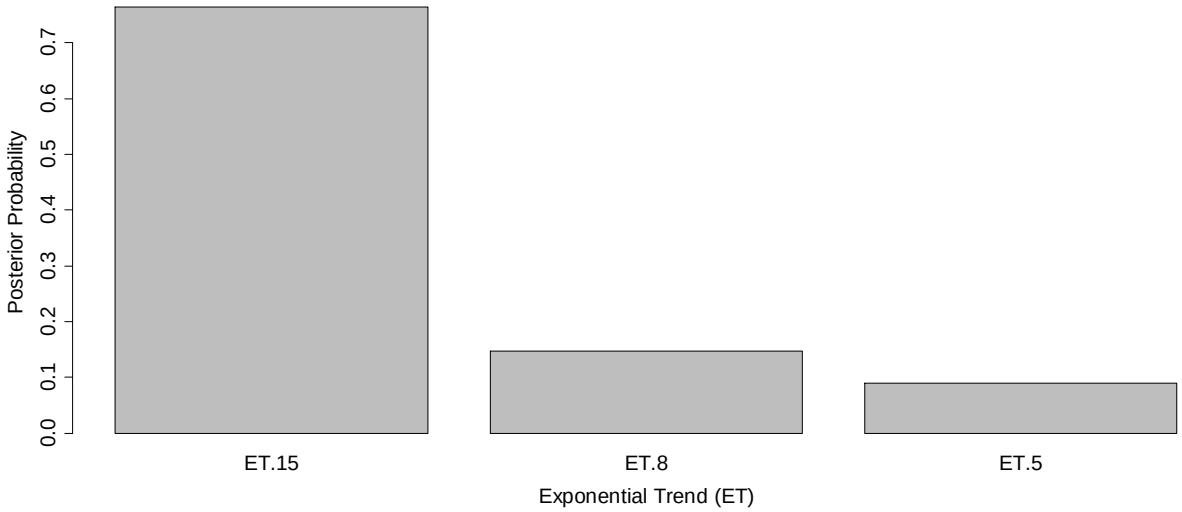


Chart 12: Frequency, Growth Rates, 1996–2009

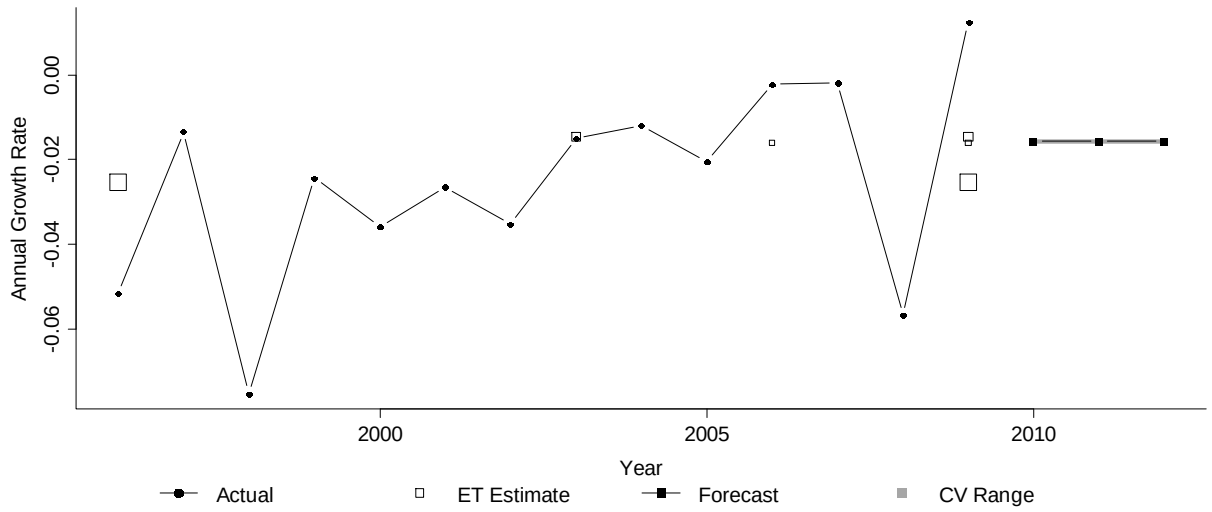
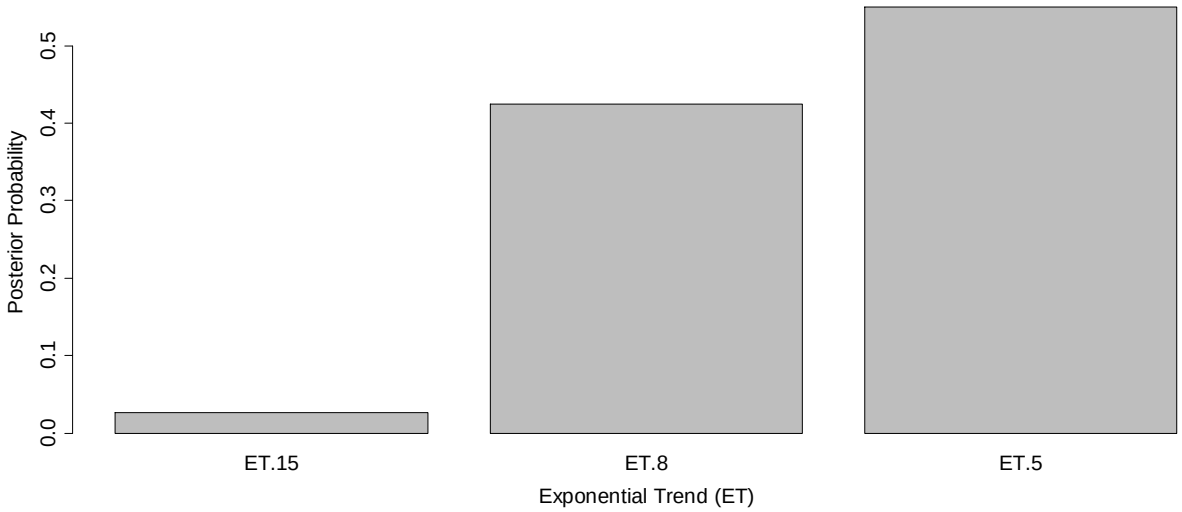


Chart 13: Frequency, Posterior ET Probabilities



Charts 14 and 15 display the findings for the severities. As mentioned, the ET estimates and the BTS forecast for the severities are backed out of the CAGR values of frequency and the pertinent loss ratio. Chart 14 reveals the upward drift in indemnity severity that contributes to the drift in the

severity loss ratio. Chart 15, on the other hand, indicates that the growth rate of the medical loss ratio is highly stationary

Chart 14: Indemnity Severity, Growth Rates, 1996–2009

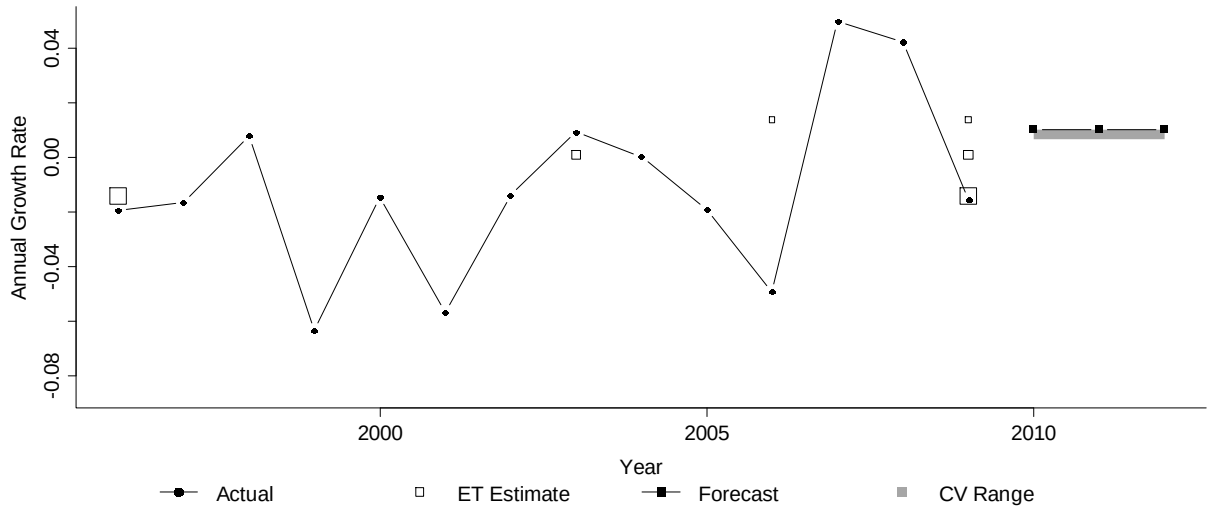
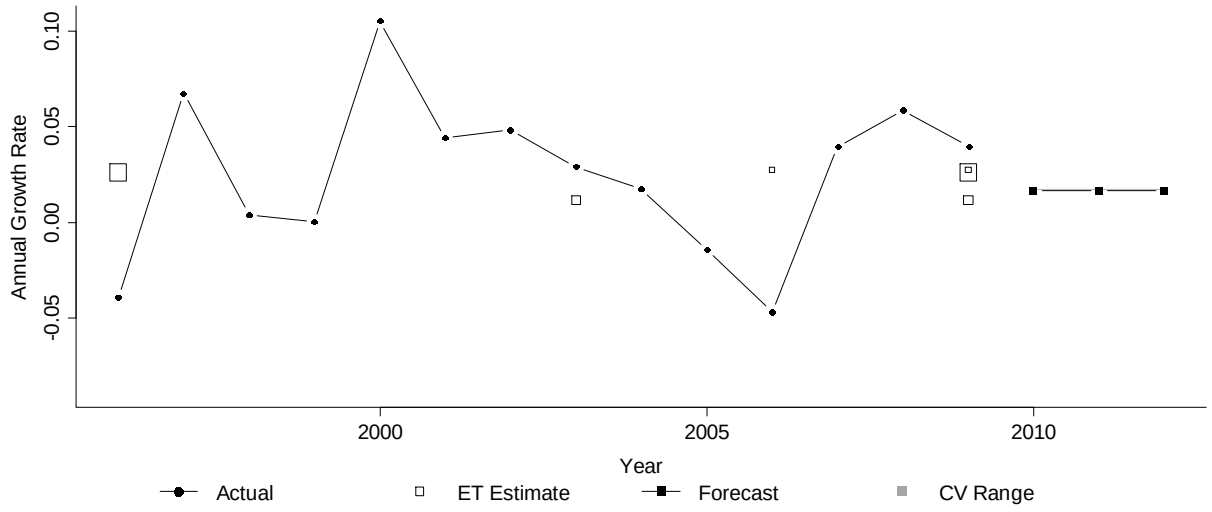


Chart 15: Medical Severity, Growth Rates, 1996–2009



7. CONCLUSION

The BTS is an attempt to formalize the trend selection process in NCCI aggregate ratemaking. The information processed by the BTS is confined to past forecast errors, which represents both a strength and a weakness of this model. On one hand, the BTS is not prone to biases in human decision making. On the other hand, the model is not capable of processing information that is not incorporated in the data, such as changes in the economic or legislative environments that occurred between the end of the experience period and the time of decision making.

Given the relative shortness of the loss ratio data series available in NCCI aggregate ratemaking, and considering the high degree of volatility of these series, the use of covariates bears the risk of fitting to noise; such overfitting can cause considerable forecast errors. Further, economic conditions tend to mean-revert; for instance, economic recessions usually last less than a year. The comparatively long trend period of slightly more than three years typically suffices for average economic conditions to reestablish themselves. Thus, mean reversion limits the loss of information associated with the absence of macroeconomic covariates as long as the decision maker relies on long-term averages.

A major strength of the BTS is its robustness. Although the BTS underperforms the 15-point ET on recent ratemaking data sets, the BTS performs well in an environment where the nature of the data-generating process is changing. Because there is always a degree of uncertainty surrounding the process that generates the growth rates of the loss ratios (or alternatively, frequency and the severities), the BTS emerges as a robust decision rule.

The strength of the BTS becomes apparent when, for instance, in a given state, the growth rate of frequency repeatedly falls short of its long-term average. Such was the case during the boom in the housing market in the early 2000s, as documented in Schmid [6]. On one hand, the 15-point ET generates the smallest forecast errors in general; on the other hand, observing the frequency rate of growth falling short of trend for many years running poses a challenge to the decision maker. Although these frequency growth rates ultimately mean-reverted, the time it takes for the trend to reestablish itself in the observed rates of growth may be longer than a decision maker is prepared to tolerate.

8. APPENDIX

8.1 JAGS Code (BTS)

```
## Bayesian Trend Selection (BTS)
model
{
  for(series.i in 1:n.series){
    for(year.i in 1:n.model.years){
      actual[series.i,year.i] ~ ddexp(mu[series.i,year.i],tau[series.i])
      mu[series.i,year.i] <- exp.trend[series.i,year.i,model.id[series.i]]
    }

    tau[series.i] ~ dgamma(0.001,0.001)
    sigma[series.i] <- sqrt(2)/tau[series.i]

    model.id[series.i] ~ dcat(model.id.p[series.i,1:n.exp.trend])
    model.id.p[series.i,1:n.exp.trend] ~ ddirch(lambda[series.i,])
  }
}
```

8.2 BLS Manufacturing Injury and Illness Incidence Rate, 1926–2010

The long series of manufacturing injury and illness rates (1926–2010) is inspired by research at the Federal Reserve Bank of Dallas. In its Annual Report, authored by Michael Cox and Richard Alm (dallasfed.org/assets/documents/fed/annual/2000/ar00.pdf), the Bank published a series of injury rates per 1,000 full-time workers in manufacturing for the period 1926 through 1999 (page 8).

The Dallas Fed series of manufacturing injury and illness incidence rates runs from 1926 through 1999 and draws on four sources:

- *Historical Statistics of the United States: Colonial Times to 1970*, Census Bureau, 1975, www.census.gov/prod/www/abs/statab.html; incidence rates are available from 1926 through 1970
- *Statistical Abstract of the United States*, various years, www.census.gov/prod/www/abs/statab.html
- Webster [11]; incidence rates are available from 1977 through 1997
- BLS

From 1926 through 1970, the rate is the average number of disabling workplace injuries per million man hours worked; effective 1958, the industry definition was revised to conform to the 1957 edition of the Standard Industrial Classification (SIC) Manual. (The SIC was developed in the 1930s; www.census.gov/epcd/www/sichist.htm.)

Bayesian Trend Selection

From 1972 on, the number of injuries and illnesses per 100 full-time workers is computed as (N/EH) × 200,000, where N is the number of injuries and illnesses and EH is the total hours worked by all employees during the calendar year; 200,000 is the base for 100 equivalent full-time workers (working 40 hours per week, 50 weeks per year).

The *Statistical Abstract of the United States* (1976) reports the old incidence series through 1970 and the new series for 1972. The 1971 incidence rate in our data set, which equals 15.4, was obtained by means of linear interpolation, since no incidence rate is reported by the *Statistical Abstract of the United States*, issues 1972 through 1974, 1976 through 1977. The 1970 number equals 15.2, and the 1972 number equals 15.6. Although, due to methodological differences, the new series is not comparable to the old series, the structural break between the two series is insignificant. It is worthy of note that the record keeping and data reporting process for the incidence rate changed significantly with OSHA (Occupational Safety and Health Administration, www.osha.gov/) becoming operational in 1971.

Starting in 2003, the manufacturing industry is no longer defined by the Standard Industrial Classification Manual, www.osha.gov/pls/imis/sic_manual.html, but instead is based on the newly created NAICS, www.census.gov/eos/www/naics/. Prior to adopting NAICS, the definition of manufacturing had been updated periodically to conform to SIC Manual revisions.

Acknowledgment

Thanks to Eric Anderson and Sikander Nazerali for their research assistance.

9. REFERENCES

[1] Armstrong, J. Scott, "Evaluating Forecasting Methods," in: J. Scott Armstrong, ed., *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell (MA): Kluwer Academic Publishers, **2001**, 365–382.

[2] Brooks, Ward, "A Study of Changes in Frequency and Severity in Response to Changes in Statutory Workers Compensation Benefit Levels," Casualty Actuarial Society *Proceedings*, **1999**: Vol. LXXXVI, No. 164 and 165, www.casact.org/pubs/proceed/proceed99/99046.pdf.

[3] Ellison, Martin, and Thomas J. Sargent, "A Defence of the FOMC," **2010**, July 10, files.nyu.edu/ts43/public/research/fomc_tom_5.pdf.

[4] Evans, Jonathan P., and Frank Schmid, "Forecasting Workers Compensation Severities and Frequency Using the Kalman Filter," Casualty Actuarial Society *Forum*, **2007**: Winter, 43–66, www.casact.org/pubs/forum/07wforum/07w49.pdf.

[5] Hansen, Lars Peter, and Thomas J. Sargent, *Robustness*. Princeton (NJ): Princeton University Press, **2008**.

[6] Schmid, Frank, "Loss Cost Components and Industrial Structure," **2011**, National Council on Compensation Insurance, Inc., www.ncci.com/documents/Loss_Cost_Components_and_Industrial_Structure-2011.pdf.

[7] Schmid, Frank, "Statistical Trend Estimation with Application to Workers Compensation Ratemaking," Casualty Actuarial Society *Forum*, **2009**: Winter, 330–40, www.casact.org/pubs/forum/09wforum/schmid.pdf.

[8] Schmid, Frank, "Workplace Injuries and Job Flows," **2009**, National Council on Compensation Insurance, Inc., www.ncci.com/Documents/WorkplaceInjuries-0709.pdf.

[9] Taylor, Shelley E., "The Availability Bias in Social Perception and Interaction," in: Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases*. Cambridge (UK): Cambridge University Press, **1982**, 190–200.

[10] Tversky, Amos, and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185, **1974**, 1124–1131, reprinted in: Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases*. Cambridge (UK): Cambridge University Press, **1982**, 3–20.

[11] Webster, Timothy, "Work-related Injuries, Illnesses and Fatalities in Manufacturing and Construction," *Compensation and Working Conditions*, BLS, **1999**: Fall, 34–37, www.bls.gov/opub/cwc/archive/fall1999brief3.pdf.

Abbreviations and Notations

BLS	Bureau of Labor Statistics
BTS	Bayesian Trend Selection
CAGR	Compound Annual Growth Rate
CV	Cross-Validation
ET	Exponential Trend
FTE	Full-Time Equivalent
MCMC	Markov Chain Monte Carlo Simulation
NAICS	North American Industry Classification System
NBER	National Bureau of Economic Research
NCCI	National Council on Compensation Insurance, Inc.
OSHA	Occupational Safety and Health Administration
RMSE	Root Mean Squared Error
SIC	Standard Industrial Classification

Biographies of the Authors

Frank Schmid, Dr. habil., was, at the time of writing, a Director and Senior Economist at the National Council on Compensation Insurance, Inc.
Chris Laws is a Research Consultant at the National Council on Compensation Insurance, Inc.
Matthew Montero is a Senior Actuarial Analyst at the National Council on Compensation Insurance, Inc.