Forecasting and Modeling Worker Injury Rates in Ontario

Ian Moore
imoore@iwh.on.ca

Plenary Talk, January 12th, 2010

What we will be covering …

- Review
- Motivation
- Study Data Sample
- Materials and Methods
- Forecast Results and Discussion
    1. Long term growth trend
    2. Seasonal and periodic fluctuations
    3. Business cycle fluctuations
- Goodness of Fit
- Volatility of Non-Deterministic Components
- Implications and Applications
- Summary

Review

- Worker injuries are classed into two main groups: lost time injury (LTI) and no lost time injury (NLTI) counts
- Having this global distinction, equips workers compensation authorities the knowledge to provide more efficient cost effective prevention strategies in the workplace environment
- Most studies on modeling workers' compensation claim rates and the business cycle can be found in the social sciences literature (Bhattacherjee et. al. (1994); Brooker et. al. (1997); Freivalds et. al. (1990); Iyer et. al. (2005); Kuhn et. al. (1994); Shuford et. al. (2006))
- Some of the studies are not forecasting studies, but consider the relationship between markers of the business cycle (such as unemployment rates) and claim rates.

Review (2)

- Formal forecasting, from a time series perspective, attempts to study all the time series components in order to identify a reasonably good fit to make specific forecasts into the future
- There are a few studies in the broader arena of accident analysis that use time series techniques to forecast injury rates. Most use ARIMA modeling techniques
- While a large share of the variance in time series of claim rates is composed of a growth trend, most of the month-to-month fluctuation is comprised of seasonal and periodic behavior, but researchers unfamiliar with time series, can easily interpret these as random events or neglect them altogether

Review (3)

- With the time series tools we used in this study, we are able to investigate periodic behavior with greater depth and precision than with more widely used time series tools (e.g. ARIMA model)
- In this study we use a combination of standard and widely used time series methodologies
- Building on these methodologies, we introduce newer tools, such as Thomson's multitaper spectral estimator method and harmonic F-test (Thomson, (1982)), which have been widely used in various fields of study such as geophysics, climatology, space physics and digital communications, within the last 20 years, but have not been used in the social sciences

Motivation

- We investigated:
  1. Three time series forecast models to assess which model works best with time series data on workers' compensation lost time, no lost time and total claim rates
  2. Main factors that drive these rates
- The approach we use to select the best model is of value to researchers who are seeking to undertake forecasting with time series data
- A well-designed forecasting method for workers' compensation claim rates can be an effective business-planning tool for insurance authorities to assist with tracking trends and allocating resources

Study data sample

- The claim rate data for this study is sampled monthly and covers the time period from 1991 to 2007
- The numerator of these claim rates represents the worker injury counts (i.e. no lost time, lost time and total) across Ontario and was retrieved from the Workplace Safety & Insurance Board (WSIB)
- The denominator represents the total number of working employees across Ontario and was retrieved from Statistics Canada's Labor Force Survey (LFS)

Materials and Methods

- The tools and methods used in this study come from the field of time series analysis, which represents a significant branch of statistics
- Three forecast models
  1. Least squares using mean, trend and periodic components in the design matrix
  2. Seasonal autoregressive integrated moving average (SARIMA) using Kalman filter state space equations
  3. Additive version of the Holt-Winters method
- Other tools
  - Thomson's harmonic F-test to study constant periodic fluctuations
  - Stock and Watson model to study business cycle fluctuations
  - Other diagnostic tools not discussed today

Forecasting: Results and Discussion

- Classified into two parts
  A. Distribution of variance in claims compensation rates
  B. Three forecasting methodologies
- Distribution of variance:
  1. Long term growth trend
  2. Constant periodic (mainly seasonal) behaviour
  3. Business cycle behaviour
- Three forecasting methodologies:
  1. Holt-Winters
  2. SARIMA using state space projection equations
  3. Least squares projected onto a design matrix consisting of mean, trend, periodic

Forecasting: Long Term Growth Trend

- The long term growth trend is the largest contributor to the overall variance in the workers' compensation claim rate.
- This negative growth pattern is least understood where causal effects have yet to be fully investigated
- Possible driving factors
  - Automation of high hazard jobs
  - Better educated workforce
  - Legislative reforms
  - Shifting of risk to small businesses and lesser developed countries
- After surveying various large audits taken from 1992 to 1996, Conway et. al discussed why the decrease is not due to an increase in under reporting

Forecasting: Constant Periodic Fluctuations

- Not very well discussed from a quantitative perspective in the literature
- Large majority of constant periodic behaviour (i.e. deterministic non changing periodic behaviour) is seasonal, as there is other periodic behaviour that exists outside the annual seasonal cycle
- Found a significant 3-month lag in the phase of the seasonal component between total injury rate and the Ontario unemployment rate
- Injury rates experience a seasonal low in December while unemployment experience seasonal low in August
- At this point, we can not say whether this lagged correlation is causal

Forecasting: Constant Periodic Fluctuations (2)

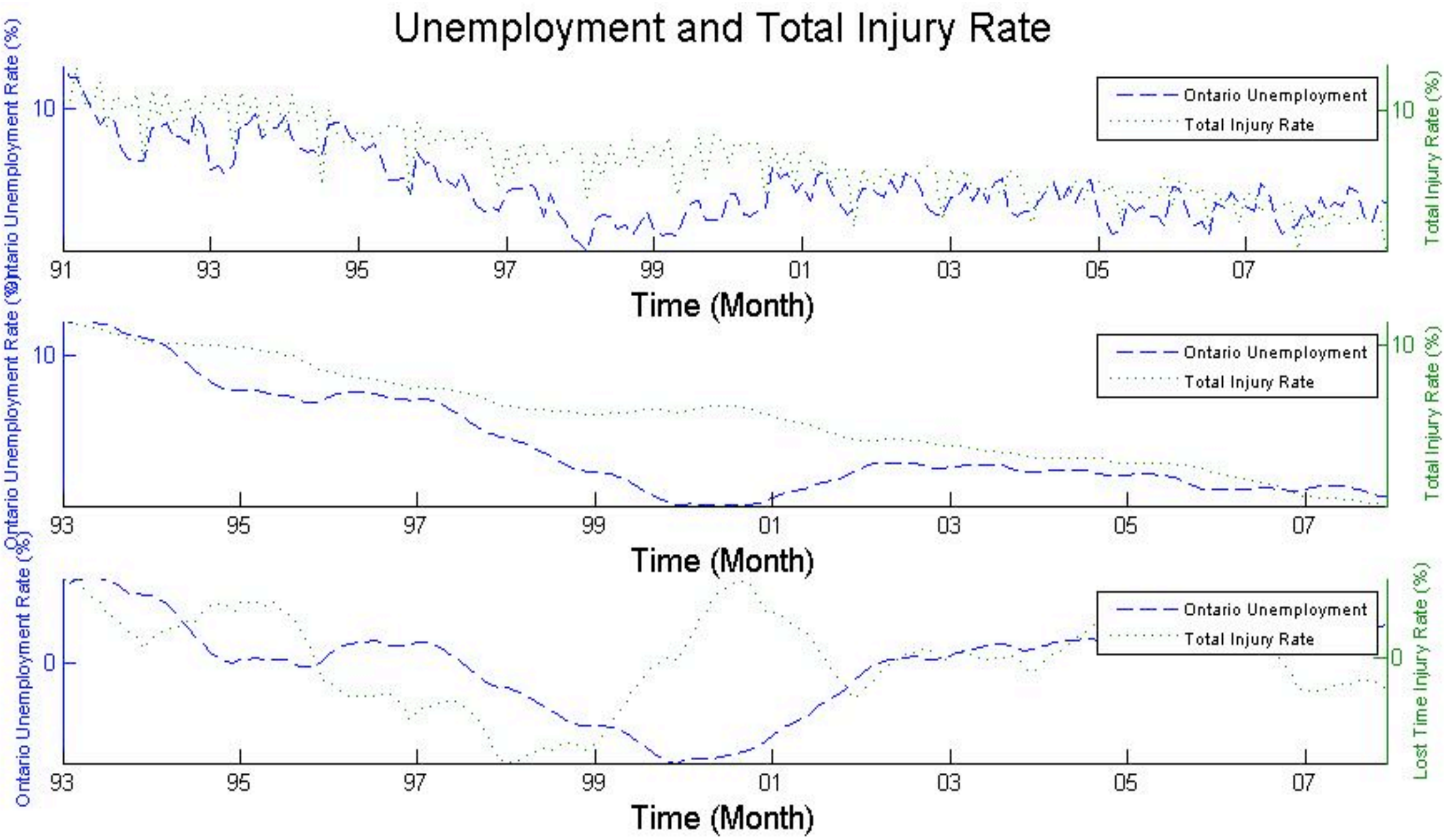
Table 1: Highly significant non-trivial frequencies derived from the harmonic F-tests.

LTI Cycle/year	Significance	%	NLTI Cycle/year	Significance	%	Total Cycle/year	Significance	%
2	0.99713	0.098501	3	0.99997	1.83E-01	4.9091	0.99854	2.16E-02
5.3824	0.99584	5.84E-03	2	0.99992	1.50E-01	2	0.99583	2.85E-01
0.85294	0.99469	6.37E-03	1	0.99957	0.2164	5	0.99308	1.01E-01
3	0.99467	1.57E-01	5	0.99909	2.65E-01	1	0.99274	0.68172
3.6176	0.99438	6.41E-03	4.1765	0.99532	1.67E-01	1.3939	0.98935	2.27E-02
3.2647	0.99217	5.03E-03	4	0.99313	0.074577	0.72727	0.98905	1.25E-01
4.1765	0.99109	8.31E-02	5.1765	0.98333	5.69E-02	2.4697	0.98665	6.11E-03
4	0.99076	8.32E-02	3.8235	0.98005	3.61E-02	1.8333	0.985	1.62E-02
5	0.98497	0.1777	1.8235	0.97836	2.54E-02	3.5909	0.98266	1.55E-03
1.1176	0.98465	0.024303	0.52941	0.97802	0.039878	4.4697	0.98223	1.97E-03
5.1765	0.98398	3.51E-02	.20588	0.9728	9.96E-03	2.2121	0.97878	3.12E-02
0.5	0.96736	2.11E-02	1.7059	0.96374	3.11E-02	0.36364	0.97666	0.23607
1.5882	0.96396	1.93E-02	2.4706	0.95736	2.23E-02	5.3939	0.97572	1.96E-02
5.5588	0.96382	5.08E-03	2.5588	0.95197	1.30E-02	1.1818	0.97308	4.02E-02
0.52941	0.95394	2.28E-02				3.6515	0.96896	1.50E-03
						2.3333	0.96587	3.29E-02
						2.6818	0.9631	4.26E-03
						1.6061	0.96128	0.024982
						1.4848	0.9591	5.61E-02
						0.21212	0.95107	2.52E-01

Forecasting: Business Cycle Fluctuations

- To study the business cycle we use a well-known methodology developed in the late 1980s by Stock and Watson
- This model utilizes an advanced smoothing operation which is an adaptation of the Kalman filter state space equations
- Originally, this model was implemented by Stock and Watson as a basis to choose a number of proxy variables, called coincidence variables, to mimic the behavior of Gross Domestic Product and to study international business cycle dynamics
- Assuming the same initial statistical assumptions, we have found this model useful to study business cycle fluctuations in workers' compensation claim rates

Forecasting: Business Cycle Fluctuations (2)



Institute for Work & Health | Research Excellence Advancing Employee Health

Forecasting: Distribution of variance

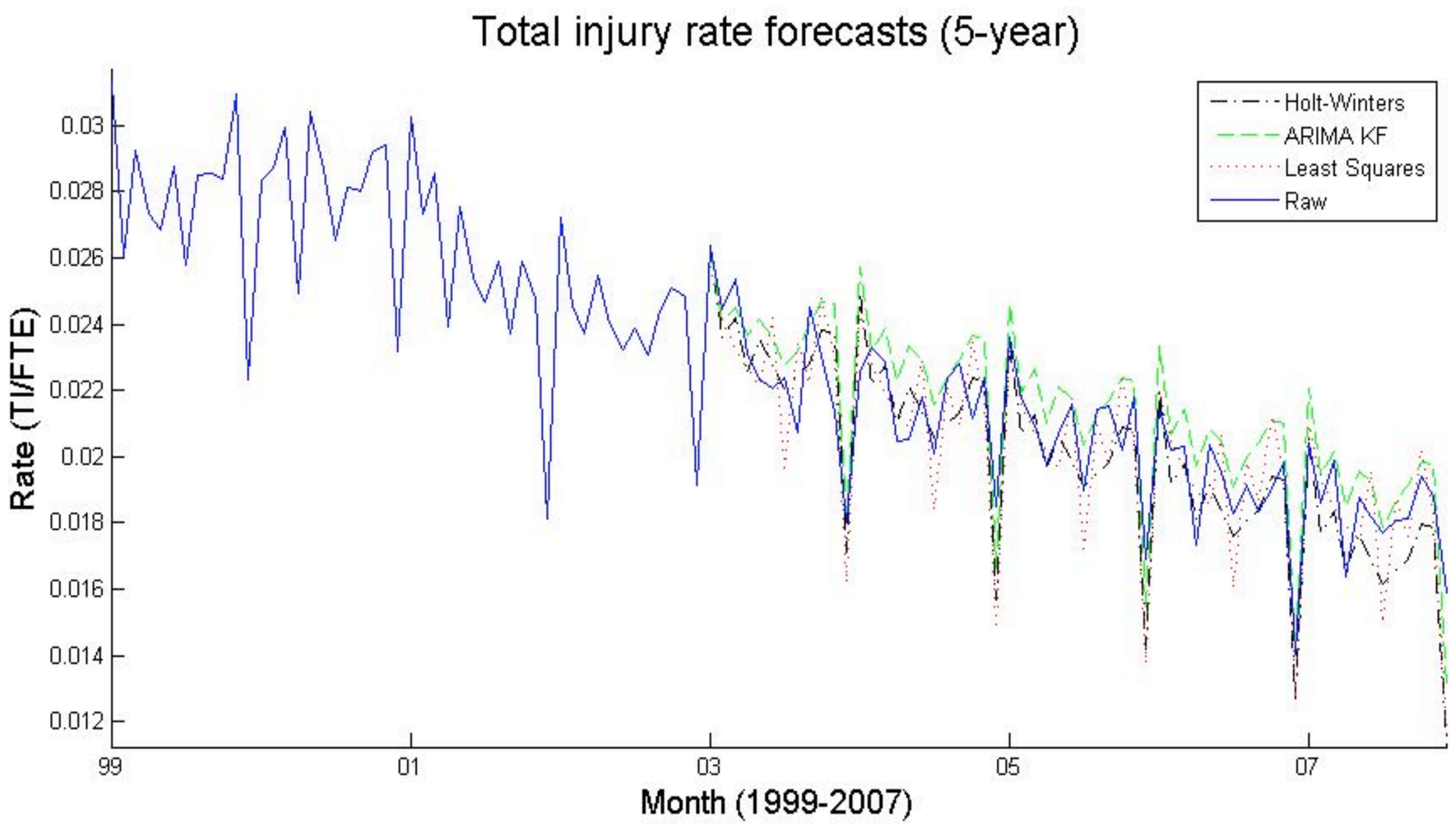
Table 2: Confidence intervals of variance explained (in order of relevance) found in claims rates: ($\hat{\ell}_t$) growth trend; (v_t) measurement error; ($\hat{\gamma}_t$) business cycle fluctuations. The measurement error, v_t, as described in the Stock and Watson single index model, was explained predominately by the variance of the periodic nature found within these rates, as validated by observing the relative flatness of the spectrum estimates.

	LTI ($\hat{\sigma}^2$)		NLTI ($\hat{\sigma}^2$)		TI ($\hat{\sigma}^2$)	
	5%	95%	5%	95%	5%	95%
$\hat{\ell}_t$	4.073E-5	5.646E-5	2.503E-5	3.470E-5	1.324E-4	1.835E-4
v_t	7.103E-6	1.072E-5	1.072E-5	1.486E-5	2.745E-5	3.811E-5
$\hat{\gamma}_t$	1.660E-6	1.198E-6	1.349E-6	1.870E-6	2.846E-6	3.944E-6

Forecasting: Driving Factors

- Knowing the factors driving workers' compensation claim rates is useful for analysts and policy makers to use as a benchmark for their models and decision making practices
- To consolidate these observations with previous work, according to contribution in variance, we conjecture that these factors are driven (in order of relevance)
  1. Growth trend
  2. Seasonal cycle
  3. Business cycle
- There are most likely other factors (e.g. impulse shocks due to legislative reforms)
- We also conjecture that these components can be sub categorized, possibly using a multivariate approach

Forecasting: 5-year Monthly Time Series Projections



Institute for Work & Health | Research Excellence Advancing Employee Health

Forecasting: Comparative goodness of fit

Table 3: Comparative evaluation of five-year forecasts (Jan. 2003 to Dec. 2007) using historical information (Jan. 1991 to Dec. 2002) of worker LTI, NLTI, and TI rates.

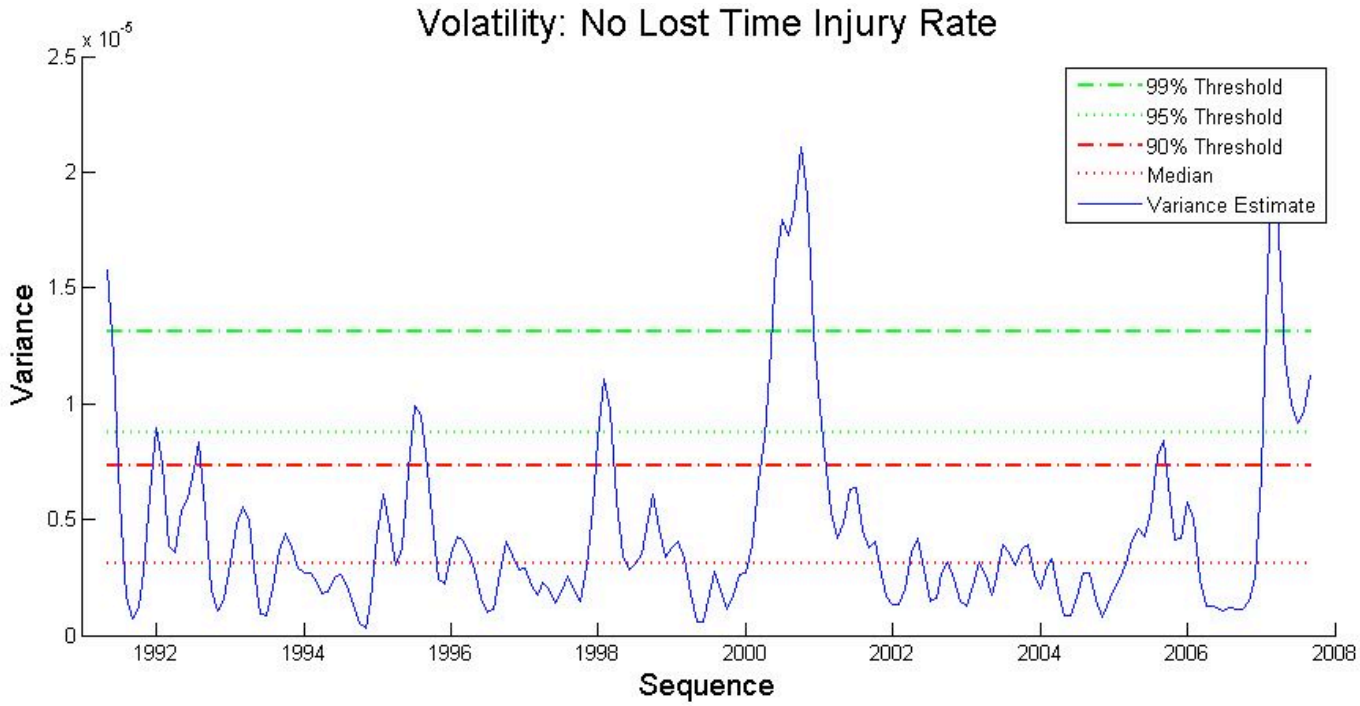
		Adj. R^2	$\hat{\sigma}$ (5 %)	$\hat{\sigma}$ (95 %)
Least Squares	LTI	0.721***	2.11215E-06	2.85813E-06
	NLTI	0.636***	6.43790E-06	8.71168E-06
	TI	0.710***	1.22895E-05	1.66299E-05
SARIMA-KF	LTI	0.817***	1.12547E-06	1.52298E-06
	NLTI	0.681***	5.57834E-06	7.54853E-06
	TI	0.787***	8.81657E-06	1.19305E-05
Holt-Winters	LTI	0.838***	1.33038E-06	1.80025E-06
	NLTI	0.719***	4.9115E-06	6.64617E-06
	TI	0.813***	8.09091E-06	1.09485E-05

*** $p < 0.0001$ (H0: no correlation when considering estimated correlation coefficient)

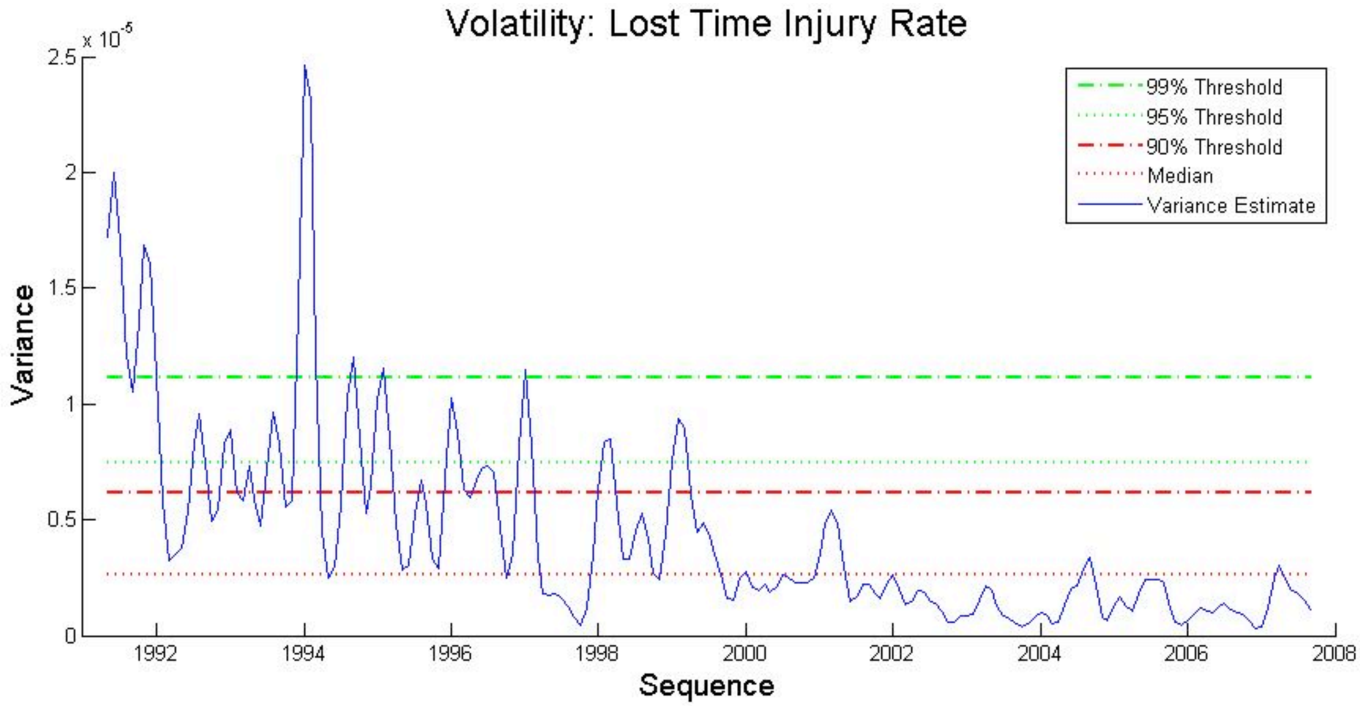
Forecasting: Comparative goodness of fit (2)

- Based on observations of the R^2 estimates and residual MSE estimates from previous table (and other diagnostic tools not discussed today) we have determined the order of performance of the aforementioned models
  1. Holt-Winters
  2. SARIMA using state space projection equations
  3. Least squares projected onto a design matrix consisting of mean, trend, periodic
- Holt-Winters approach is also considered to be ad-hoc over the ARIMA approach (Harvey, 1993)
- From a time series forecasting standpoint, when considering all factors, many consider the Holt-Winters method to be the better choice over the ARIMA model (Chatfield, 2004)

Volatility of LS model residual: No Lost Time Injuries



Volatility of LS model residual: Lost Time Injuries



Limitations

- To maintain consistency in this study, we used only univariate time series forecast approaches
- There is some debate in the literature as to which approach, univariate versus multivariate, is better (Chatfield, 2004)
- Given the good performance of the univariate forecast models used in this study, which explained between 70 to 80 percent of the variance, it is not clear whether a multivariate approach would make significant improvements
- A multivariate approach may however shed more light on factors that drive these rates

Retrospective Implications and Applications

- The approach we use is of value to others who are seeking to select a model for time series forecasting of workers' compensation claims rates
- Methods can be implemented on workers' compensation claims rates stratified by industry sector, gender, age etc.
- Useful to large worker compensation authorities wishing to make long and short range business planning decisions
- Can easily implement these tools into a ready-to-use and tested software tool kit

Summary

- Motivated by the need to develop long range forecasts
- Data used came from WSIB and LFS
- Tested the Least Squares, SARIMA and Holt-Winters approach and found Holt-Winters approach to provide the most accurate forecast
- Three *main* components driving claims rates
  1. Long term growth trend
  2. Constant periodic fluctuations
  3. Business cycle fluctuations
- Volatility of non-deterministic components are falling more rapidly with lost time injury rates over no lost time injury rates from 1991 to 2008
- Forecasts are useful to large workers compensation authorities for long range business planning

Institute for Work & Health

Research Excellence
Advancing Employee Health

www.iwh.on.ca