New Methods for Forecasting the Cox Process

By

Kenneth Matthew Hutchison

A Project Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

In Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Major Subject: INDUSTRIAL AND SYSTEMS ENGINEERING

Approved:

____

Thomas Reed Willemain, Thesis Adviser

____

Wai Kin Chan, Thesis Adviser

____

Jennifer K. Ryan, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York
November, 2011
(For Graduation May 2012)

ABSTRACT

Gathering and analyzing data is becoming increasingly pertinent and complex in both research and enterprise. The increase in data collection suggests that Cox processes (Poisson processes with nondeterministic mean) may be more noticeable than in the past. Decisions in enterprise and financial transactions are often automated and executed on the order of milliseconds. It follows methods are needed to effectively grapple with large masses of data and to do so in a reasonable time frame. Current methods used in the analysis of the Cox process, such as principal component analysis, Markov chain Monte Carlo methods and maximum likelihood estimation are limited in that the problems must be reasonably small to solve (because of the memory storage requirements) and computation of the estimates may take hours or weeks.

This thesis is concerned with developing and demonstrating algorithms to improve the accuracy and speed of estimating the parameters of a Cox process. Estimating the intensity function of the doubly stochastic Poisson (Cox) process is not trivial and has plagued researchers for decades. In this thesis, piecewise approximation of the intensity and empirical algorithms based on exponential smoothing are used in order to estimate and forecast a Cox process with stochastic intensity function given by a class of time series models.

The underlying distribution of the intensity process will be generally modeled as an Autoregressive Integrated Moving Average process of type ARIMA(1,0,0). A modified reversed exponential smoothing method is proposed as a way to approximate the intensity process. This thesis develops a model for forecasting processes by developing a method to estimate the parameters of the underlying process as an AR(1) model. This model is chosen because it is reasonable in the case of time series applications with sparse counts to suspect that an AR model is driving a Cox process. Also, for series where data is lacking, the estimation of models more complex than ARIMA(1,0,0) may be inappropriate.

This thesis examines the accuracy of the estimators for lead time demand distribution forecasting and backcasting using synthetic data in a factorial computational experiment. The estimators are also assessed by forecasting 3 corporate datasets with 1000 items, 1 corporate dataset with 772 items, and 1 corporate dataset with 35 items. On the actual data, the estimation methods were found to increase forecast accuracy up to 40% relative to a benchmark method given by a simple Poisson model.