

Xiaojuan Zhu

June 12, 2015

Abstract

1 Introduction

Employee turnover is a topic that has drawn the attention of management researchers and practitioners for decades, because employee turnover is both costly and disruptive to the functioning of most organizations (Staw, 1980; Mueller and Price, 1989; Kacmar et al., 2006), and both private firms and governments spend billions of dollars every year managing the issue according to Leonard (2001). Therefore, understanding the causes of turnover: retirement and voluntary quit, examining the internal and external impacts, effectively forecasting the turnover by these two causes, and measuring the effectiveness and to what extent of the HR policy at firm and departmental levels are the key questions in this study for reducing it and for effective planning, budgeting, and recruiting in the human resource field. As a funded research project, a large organizational secondary dataset including 12-year employees demographic information and records is transformed, analyzed and modeled by Cox proportional hazard regression models using competing risks analysis to examine the statistically significant factors and to predict employees' conditional retiring and voluntary quitting probabilities. The dataset are also employed to logistic regression and time series models for compare the performance of cox proportional hazard model. This study also examines the forecasting capability of Cox proportional hazard model on the data with two kinds of bias (left truncation and right censor) by simulation.

2 Literature Review

3 Data Preparation

objective is to forecast employee retirement and voluntarily quit using statistical model

1. data description

a. The data is employee demographic information and records windows 10 years, data's detailed information.

b. data series plot and describe 2008 event.

2. economic indicator variables, sources and what they are, why you select these economic index

4 Model Development and Description

4.1 How to model employee turnover

4.1.1 Cox ph regression model

1. what is cox ph regression model. why not parametric. baseline cannot fit to any parametric model
2. cox regression without/ with time dependent variable
3. Forecast employee turnover c. how to forecast and calculate employee turnover number
4. conditional probability (The prediction is conditional probability): given the employee is survival at last year, what is the probability they survival or quit for this year.
5. the total number of employee turnover is the aggregate all turnover probabilities of employees at each year

4.1.2 Two data bias: right censor and left truncation

what is right censor, how to deal with right censor
what is left truncation how to deal with left truncation
simulation study.

4.1.3 Time dependent covariate

I. due to 2008 intervention, each employee have different age at 2008, we have use counting process. Each employee has up to 3 records related to the time, which is before 2008, in-between 2008, and after 2008. The variables related to time are changed based on years, such as policy, age, year of service. Policy is a dummy variable in the form as shown in equations. Two age points are calculated for each record at each period: one is age at beginning of the certain period, named "age at start"; and the other one is age at end of the curtain period, named "age at end". Two year of services variables are also generated for each record at each period: one is year of service at the beginning at the period, named "YCS at start"; the other one is the year of service at the end of the period, named "YCS at end".

II. economic indiecaters, is time dependent external variables, we have split each employee's the record into yearly records, so each employee has up to 12 records. What is counting process.

4.1.4 Competing risks

what is competing risks. competing ricks can help forecasting employee retirement and voluntary quit.
why select these two reasons to model.

4.1.5 Stratification model

- I. what is stratification model;
- II. how to select a stratify variable

4.1.6 Variable selection

all the variables are putting into model and selected by backwards selection methods based on P-value.

4.2 Model validation and comparison

4.2.1 Measurements

AIC, BIC, MAPE, c statistic.

4.2.2 Model validation

split data into two ways: both training and holdout, training to build the model and holdout is to validate the model, and compare the actual vs forecasting, calculated MAPE and C statistics.

WAY1: training (10 years)and holdout (2 years). forecasting the total number of employee leaving the organization, compared to the actual number to calculate MAPE and C statistics. compare to logistic regression and time series methods to shown survival is better.

WAY2: random split the data into training (2/3) and holdout (1/3), using c statistics to select one best survival model.

4.3 Results

4.3.1 Right censor and left truncation simulation results

4.3.2 Retirement model without external variables

1. four survival model have been generated for comparison.
2. significant variables.
3. stratification variable determination.
4. model comparison based on validation way 1 and way2 (one table show all the models)

4.3.3 Retirement model with external variables

best model and tested which variable does significantly impact on retirement.

4.3.4 Voluntary quit model without external variables

I. dependent variable are YCSH, because age is not able to predict well. II. shorten the length of risk set. iii. model comparison. (survival model, time series model, and logistic regression model)

4.3.5 Voluntary quit model with external variables

i. tested which variable does significantly impact on employee voluntary quit.

5 Conclusions and Managerial Implications

References

- K. M. Kacmar, M. C. Andrews, D. L. Van Rooy, R. C. Steilberg, and S. Cerrone. Sure everyone can be replaced but at what cost? turnover as a predictor of unit-level performance. *Academy of Management Journal*, 49(1):133–144, 2006.
- B. Leonard. Turnover at the top, May 2001.
- C. W. Mueller and J. L. Price. Some consequences of turnover: A work unit analysis. *Human Relations*, 42(5):389–402, 1989.
- B. M. Staw. The consequences of turnover. *Journal of Occupational Behaviour*, pages 253–273, 1980.