COMPARATIVE ANALYSIS OF MATHEMATICAL MODELS IN EPIDEMIOLOGY- THEIR ADAPTATION TO EXPLAIN HERDING IN FINANCIAL MARKETS

Saurabh Tomar Assistant Professor and Research Scholar Bhlai Institute of Technology,Durg,C.G,
Daljeet Singh Wadhwa Associate Professor and Research Supervisor Bhlai Institute of Technology,Durg,C.G,India : tomar.mba@gmail.com ; daljeetsingh.bit@gmail.com

Abstract

There are different approaches in the use of a mathematical model to explain the spread of infectious diseases. Epidemiological models like Cox, B Spline, SIR, and Transmission probability are used to explain the herding effect. The bulk of the studies shows the use of these models in Epidemiological research. This paper tries to find the rationale for choosing particular models. The collected data sample could be left truncated, right truncated, interval censored and similarly, the approximation required to fit the dataset in models could be different. This paper tries to find out the appropriate model which can be used to describe herding phenomena in the case of the financial market. One of the key findings is that it is better to use the B spline model due to the presence of left truncated data in a stock market crash or a bank run.

Key Words: Herding Behavior, Epidemiological Analysis, Cox Model, B Spline Model, Transmission Probability, SIR Model

Extended Summary:

Herding is an phenomena that do explain the anomalies of Financial Market Asset pricing, to some extent. Present study try to explore different models that can be used in explaining herding phenomena. The models that were compared in the study was Cox, B Spline, SIR, and Transmission probability. These models are widely used in Epidemiological analysis in explain the infectious effect of communicable diseases. Initially the study examine the analogy between SIR model of epidemic spread and information flow during a major financial event. It divides the participant into three compartments those who are susceptible of information diffusion those who are infected and those who are recovered from the effect of information. In another comparison COX regression model is used to identify the effect of several variable at a single instance. The analysis of model was also helpful in getting the transmission probabilities. This analysis was helpful in zeroing down the factors which are to be considered while selecting a particular model to define herding in Financial Markets

Introduction

Do Individual investors "flock together" (or "herd"), when they trade securities or when they deal in another type of Financial Business, like operating a bank account? Do some investors follow the lead of others when they trade? Does a person take part in the bank run just due to information diffusion? And if they do so, then, is it possible to model the progress of this behavior by the way of the mathematical model which is used to describe the flow of epidemic. These questions have interested researchers for some time and are at the center of this paper. The basic problem in selecting a model for a particular analysis is the presence of abnormalities in the collected data sample. Data could be right-censored, left truncated, or interval-censored. Some approximation may be required for the probability density function. The effect of covariate may also be present. Social scientists' event-history data are rarely complete. Frequently, an event history is right-censored, meaning that the length of the survival time is observed to be greater than a certain value. But the precise length is unknown. Statistical procedures for right-censored data have been developed and used routinely (Cox and Oakes 1984; Tuma and Hannan 1984; Allison 1984; Yamaguchi 1991). Sometimes, event-history data are left-truncated, meaning that a

ISSN: 2278-4632
Vol-12 Issue-07 No.01 July 2022

subject has been exposed to the risk of an event for a while when it comes under observation; the length of exposure before observation may or may not be known.

The first section of this paper describes different models used in Epidemiology to explain the epidemic. Like SIR Model, Cox Model, and B Spline model...Then the second section discusses the use of the B Spline model in modeling the flow of the Aids Epidemic among the Injecting Drug user in Bangkok. The Paper also refers to the rationality of using B Spline instead of the Cox Model in Bangkok research. At last, I compare the use of the B-spline and Cox model in the study for herding in financial markets.

As described by *A. Devenow, 1. Welch / European Economic Review 40 (1996) 603-615* herding can be defined as a behavior pattern that is correlated among individuals. The coordinating mechanism required for herding can be based on some signal or based on observing other decision-makers. Herding behavior demonstrates how human and other entities act together as an individual in a group without any planned direction. There has been a wide variety of research to give some scientific bases to this kind of behavior. Be it in the biology field e.g. spread of infectious diseases or stock market crash and bubble or say in a bank run. Herding behavior can explain the judgment formation and decision-making in this type of scenario. Psychological and economic research has identified herd behavior in humans to explain the phenomena of large numbers of people acting in the same way at the same time.

To select an appropriate model to describe herding behavior in the Financial Market one has to consider several factors. The model should be able to incorporate the censuring and truncation of collected data. The covariate effect which influences herding behavior is also to be taken into consideration. Ideally, the model which is used to analyze this behavior in the financial market should be free from the approximation of probability and cumulative density function of the random variable used.

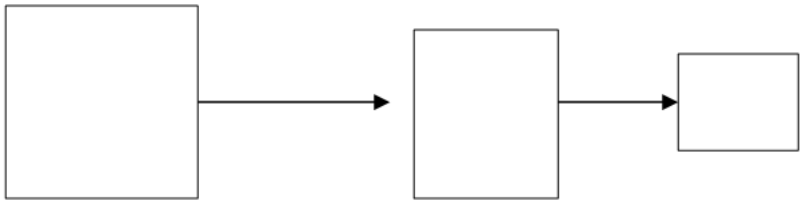
There have been different research that uses different models to predict the flow of epidemics. Some of these models are described below. We will try to evaluate each of the below-described models on the parameters which are suitable to analyze herding in financial markets

Section - 1

1.SIR MODEL:

A collection of individuals or populations can be thought of as entities having different characteristics in different fields. In this scenario, if an epidemic strikes then to understand it scientifically, different theories and statistical models have been developed. One of the most popular models is the SIR model. It is based on dividing the individual into three compartments those who are susceptible (S) to getting a disease, those who are infected (I), and those who are immune to the disease or have recovered from its effect (R).The notation denotes the number of people in each compartment at any particular instance. As time progresses there is an infusion of individuals from one compartment to another so S, I, and R can be more precise by denoting them to be the function of time (t) say by S(t), I(t), and R(t).

This dynamic part can be explained by considering an instance of an epidemic, say, the number of individuals infected from the disease increase to a particular extent called threshold then, it can be called an epidemic, now as time progresses more individuals pass from susceptible stage to infected stage and soon to the recovered stage so the flow is from S to R as time progresses. And as soon as 'I' decrease to a particular extent the disease can no longer be called an epidemic.



1.1Herding Among Investor: Explanation based on SIR Model

At any given time during a bull run or a crash in a stock market, we want to know the number of people who are infected by the decision of other investors. We also want to know the number who have been infected and have recovered, because these people now become immune to further influence. If we ignore movement into and out of the infected area, then the remainder of the population is still susceptible to herd behavior. Thus, at any time, the fixed total population may be divided into three distinct groups:

- those who are infected,
- those who have recovered, and
- those who are still susceptible

In this discussion, we will treat an Infected Individual as the one who decides to invest after coming in contact with another infected investor. Similarly susceptible individuals are the ones who take part in Financial Markets. The participant in Financial Markets will be taken as constant.

We are interested in the spread of infectious information among the participant in the financial market. The participant is divided into three compartments individuals which are susceptible to the information, those who are currently infected with the herding effect, or those who are called a recovered individual who has invested after getting infected or have taken an exit from the market and their action have been taken care of in setting up the market price. Thus we have three groups or states in which we can place individuals. In addition, we see that our data is a time series where we have several infected individuals at each point in time. Similarly, we also have several susceptible and recovered individuals at each point in time. Let us begin with some notation.

S_t = the number of susceptible individuals in the population at time *t*.
I_t = the number of infected individuals in the population at time *t*.
R_t = the number of recovered individuals in the population at time *t*.
N = the population size. (Market participants)
Similarly, we define each group as the fraction of the population as
$s_t = St/N$ (the susceptible fraction of the population at time *t*.)
$i_t = It/N$ (the infected fraction of the population at time *t*.)
$r_t = Rt/N$ (the recovered fraction of the population at time *t*.)
Since we have divided all market participants into these three groups, we have
$S_t + I_t + R_t = N$ *and*
$s_t + i_t + r_t = 1$.

1.2 Dynamics of Herd behavior:

In the process of information flow, we have flow of individuals from susceptible to the infected group and then to the recovered group. Whereas, the transmission of the infection occurs when an infected person comes into contact with a susceptible individual. Suppose on average each infected person contacts 'λ' individuals in each period. Now each contact may not result in the transmission of the herd effect. Perhaps only '*α'* percent of the contacts results in transmission. Thus the potential number of transmissions may at most be-

$\beta = \lambda * \alpha$

Where β is the average number of transmissions possible from a given infected person in each period.

If we assume that individuals are mixed randomly then each potential transmission may be from an infected person to a susceptible person which results in a newly infected person. Or a transmission may occur from an infected person to another infected person which results in nothing happening since the person is already infected. Or the potential transmission may occur from an infected person to a recovered person. Again in this case nothing changes. Since only "s_t" percent of the population is susceptible each infected person generates only $\beta * s_t$ new infections each period. Each infected person recovers at some rate. Let the fraction of the infected group that recovers or has acted be "κ "

We are now ready to describe the SIR process. Given the current state of the population in period '*t'* described by S_t; I_t and R_t we can write a series of equations that describe the motion of the system. First, let's describe the susceptible population. We begin period *t* with S_t Individuals in the susceptible population. From this population, we lose on average $\beta^* S_t * I_t$ from the population. Thus in period $t + 1$ we have:

$S_t+1 = S_t – \beta S_t I_t$ (1)

Through similar reasoning we see that:

$R_t+1 = R_t + \kappa^* \cdot I_t$ (2)

And

$I_t +1 = I_t +\beta^* s_t * I_t – \kappa * I_t = I_t (1 +\beta s_t - \kappa)$ (3)

Similarly, we could write each of these in terms of the population fractions:

$s_t+1 = s_t – \beta s_t i_t$ (4)

Through similar reasoning we see that:

$r_t+1 = rt + \kappa it$ (5)

And

$i_t+1 = i_t (1 +\beta_{st} - \kappa)$ (6)

If we add up these equations we will find that

$st+1 + i_t +1 + rt+1 = st + i_t + rt = 1.$

Let $\rho_t = 1 +\beta_{st} - \kappa$

This is the epidemic threshold for the SIR model with a constant population (we are assuming market participants to remain constant during the analysis). If ρ_t is greater than 1 then we are multiplying i_t by a number greater than 1 so $i_t+1 > i_t$. The number of infected individuals is increasing. But if ρ_t is less than 1 we are multiplying i_t by a number less than 1 so $i_t+1 < i_t$. The number of infected individuals is decreasing.

Now let us take a brief look at s_t and r_t. We know that s_t is decreasing for $i_t > 0$. And we also know that i_t equals 0 in a steady state. So, from this, we can deduce that s_t will also reach a steady-state value since it is constant if i_t is equal to 0. Now, r_t is increasing for $i_t > 0$ and is at a steady state when $i_t = 0$. Likewise in the process of time s_t, r_t will also reach a steady-state value. So if we have statistical data of initial and final condition of s_t, i_t *and* r_t, we can plot these with different values of β_{st}, and κ, and subsequently transmission probability ρ_t can be plotted.

1.3 COX MODEL

Here I will describe Cox Model by giving a case study of clinical analysis. The Cox model is based on a modeling approach to the analysis of survival data. The purpose of this model is to simultaneously explore the effects of several variables on survival. In the case of a clinical trial, the use of the Cox model allows us to isolate the effect of several independent variables. Survival time is referred to as the development of the disease symptoms up to the time of death. Survival time will be censored as there will be some observation that will not produce the desired result (Death) between the times of observation, as there will be some patients who are still alive at the time of completing the study.

The regression method introduced by Cox is used to investigate several variables at one time. It is known as proportional hazard regression analysis.

In Cox Model, the hazard function is the probability that an individual will experience an event (for example, death) within a small time interval, given that the individual has survived up to the beginning of the interval. It can therefore be interpreted as the risk of dying at time *t*.

The proportional hazard model is not based on any assumptions concerning the nature or shape of the underlying survival distribution. The model assumes that the underlying hazard rate is a function of the independent variables (covariates); no assumptions are made about the nature or shape of the hazard

function or of the residual left after regression. Thus, in a sense, Cox's regression model may be considered to be a nonparametric method. The model may be written as:

$h\{(t), (z_1, z_2, ..., z_m)\} = h_0(t)*\exp(b_1*z_1 + ... + b_m*z_m)$

Where *h (t,...)* denotes the resultant hazard, given the values of the covariates for the respective case (z_1, z_2, ..., z_m) and the respective survival time (t). The term $h_0(t)$ is called the *baseline hazard*; it is the hazard for the respective individual when all independent variable values are equal to zero. We can simplify this model by dividing both sides of the equation by $h_0(t)$ and then taking the natural logarithm of both sides:

$\log[h\{(t), (z...)\}/h_0(t)] = b_1*z_1 + ... + b_m*z_m$

So this is a linear model and can be estimated with ease. In this model, we have taken two assumptions

1. That there is a multiplicative relationship between hazard function and the log-linear function of the covariates. This is called a proportionality assumption.
2. The second assumption is that there is a log-linear relationship between the independent variable and the underlying hazard function.

This model seems to take into consideration all effects which might be encountered while analyzing herding in financial markets. It does not assume any shape of the residual curve. This is also important since we know in regression we assume the shape of residual as normal and the time series in normal assumption runs through whole number line, positive and negative whereas in financial herding the time to failure i:e time when a person gets infected by information diffusion is always positive.

Another advantage of using this model is that we can take into effect the number of covariants while finding the hazard rate of the function. As different covariates like age, education, a period for which the person is involved in Financial Markets, depth of Investment in the Financial Market (Percentage of assets invested at the time of the crash), these are likely to be the covariates whose effect has to be considered, apart from age, experience with investing, education, etc. This aspect is taken care of by using Cox- Model.

This model also takes into consideration the problem of right censuring in data. Right censoring will be present in the data as it is not possible to monitor each investor up to the time he takes a decision on investing in the bull run or on taking out his investment in case of a crash.

Section – 2

Literature Review of Bangkok study & analysis of B Spline Model

We will try to explain this model by giving the literature review of the study conducted among Injecting drug users in Thailand to estimate the transmission probability of the Human Immunodeficiency virus among injecting drug users.

2.1 Literature Review:

While analyzing the statistical models to explain the transmission probability of the spread of infectious disease I focused on two papers which were published in Applied Statistic journal and American journal of epidemiology respectively. My concentration was to analyze the models used in these articles to define the transmission probability and try to compare it with the Cox Model. This was necessary to select a model which can be used to find transmission probability in case of herding behavior in Financial Markets. I was also able to define the shortcoming and usefulness of the model while doing their comparative study. The comparative study is also done to find the usefulness of the Cox Model and the reason that why the Cox Model was not used in the following studies. This comparison was beneficial in giving me the insight of recommending the B Spline model v/s Cox model in finding Transmission Probability in the case of Herding Behavior in Financial Markets.

The first article is titled "*Estimating the Transmission probability of Human Immunodeficiency virus in Injecting Drug user in Thailand" it is written by Michael G. Hudgens et al*. (This will be referred to as article 1 in my paper)
The second article is titled *"Sub Type Specific Transmission probability of Human Immunodeficiency Virus Type 1 among Injecting drug user in Bangkok, Thailand " This is written by Michael G. Hudgens, Ira M Longini et al.* (This will be referred to as article 2 in my paper).
Both of the articles are based on the cohorts of Injecting drug users (IDU's) in the Bangkok Metropolitan Administration. These are based on data collected from 1995 to 1998. Initially, all IDU's were seronegative and they were assessed for HIV after every four months.

The model used accounts for left truncation, interval censoring, and time-varying covariates. Here the transmission probability is defined as the probability that a susceptible person becomes infected from a single contact with an infectious source. Now while selecting a model it is important to select such a model which can assess covariate effect based on transmission (*Koopman et al,1991)*

2.2 Method

Cohort of IDU's consisted of 1209 IDU's initially seronegative for human immunodeficiency virus (HIV). They were followed from 1995-1998 and checked for sere conversion at every 4 months. Finally, there was 133 seroconversion.
Covariates used in this analysis were Gender, Age, jail history, frequency of needle sharing, and casual sex. These methods gave rise to competing for risk failure time data subject to the interval censoring and left truncation.
Failure time data means IDU's can be infected with a different type of sub-virus of HIV. Similarly, interval censuring means the exact time of stereo conversion was unknown i:e any time between stereo negative visit and visit after becoming sereo converted. And left truncation means that at the time of enrollment of IDU'S they were exposed to the risk of infection.

2.3 Models Used:

2.3.1 Article 1: Transmission Probability Model:

In article 1 Author used the Non-Parametric and Flexible Parametric methods for the initial exploration of data. Since the data collected is embedded with left truncation, interval censuring the maximum likelihood method used minimal assumption for the shape of residual. So these methods are used as a first step to analyze complex patterns.

Here the model used is the proportional hazard rate model, first, it is defined for a single covariate as
$\lambda_{ij}(t/c_{ij}) = c_{ij}\, p\, \pi$ for $\tau_{ij-1} < t <= \tau_{ij}$ where,
I and j are the number of individuals and the number of follow-up visits respectively.
P is the probability of infection with a single needle sharing act with an infectious IDU, and π denotes the prevalence of HIV-infected IDU.
Then, $p\,\pi$ is the baseline hazard function and $c_{i}t$ is the sharing rate over the jth interval for the I th interval
The above model can be generalized to incorporate other covariates than c_{ij}
Now in this model, they have assumed then,p is constant over time if we try to assess the information for varying transmission probability making it a random variable Basis spline model should be used, and since only p is varying not the hazard rate the class of spline model is restricted between 0 and 1.

2.3.2 Article 2

In this paper, writers assumed similar methods and models as used in Article 1, apart from the fact that this study requires the model to specifically define transmission probability of different sub-type virus B and E. Authors assume that occurrence of these sub-type viruses is mutually exclusive.

So in the nonparametric and flexible parametric model, the Cumulative density function and baseline hazard function are treated as additive. Similarly when using the transmission probability model subtype-specific transmission probability has been calculated and from that, the relative probability of infection (P_E/ P_B) is calculated

It is to be noted that in the case of Analysis in Financial Market data the data will not be left truncated. This is because any news regarding bank run or market crash will not be present before the occurrence of the event. Even we do not require the model to account for interval censoring, since the exact time of person getting affected due to herding behavior will be known. This allows us to rely more on Cox Model.

In article 1 after constructing the model to find transmission probability, which is the probability that a susceptible person becomes infected by a single needle sharing act is, estimated. While calculating this, the effect of covariates is also considered. The time depending effect is also not considered here.

Section – 3

Comparison of Cox Model vs. B Spline Model.

The Cox model is not based on any assumption regarding the nature of the survival curve. No assumption is made about the hazard function. But two important assumptions are made regarding the relationship between hazard function and log-linear covariate.

1. There exist a multiplicative relationship between hazard function and log-linear covariate, this is called proportionality assumption
2. Given two observations with the deferent value of independent variable then the ratio of hazard function for these two observations does not depend on time.

In the given article both non-parametric and parametric methods of regression are considered. In the study, they used left truncated data which means at the time of enrolment of IDU'S they were exposed to the risk of an event (sereo conversion due to ID in this case) for some time. So these subjects tend to have lower risks at a shorter duration than those in a normal sample as the high-risk subjects tend to experience the event and drop out before observation begins.

So if they use NPMLE to model then it will try to severely overestimate the commutative probability of failure in the present case of left truncation. And this MLE will be inconsistent.

So they have used condition MLE using B-spline.

The author's study requires the effect of the covariates in the study and on the base function, and the proportionality assumption, which is a condition while using Cox Model needs to be there. Since this proportionality assumption was not present in different covariates effects they were reluctant in using the Cox model in the analysis. Also, two different base hazard functions were there for B type and E type. So only if the covariate has the same effect on both the type of hazard functions then only the proportionality assumption of Cox will hold.

In all, we can say that to model the data in the presence of covariates, where the proportionality between base hazard function and covariate effect is unknown it is better to use Bspline or other conditional Maximum likelihood estimates than to use non-parametric Cox model. This should also take note that we were using left truncated data which might be there in case of using a similar model in explaining herding behavior in Financial Markets.

I would prefer to use the Cox model only if I knew that how my base function relates to the effect of the covariate. And I will use the conditional spline model to overcome the effect of left truncation if it seems to be present in financial market data, else I will use Cox Model.

Conclusion

I analyzed different models used in epidemiological literature to find the transmission probability of infectious viruses. While doing so a critical explanation was required on the logic of selecting a particular model for the study. A careful and effective failure time requires a complete understanding of the behavior of data collection. This includes the type of truncation required, the censuring mechanism, and the number of covariates that are to be used in analyses. Based on this parameter usefulness of a particular model for the failure time analysis is based.

It is also important to know the approximation required for hazard function this helps in determining that whether a semi-parametric, Flexi parametric, or nonparametric model is to be used. One should also know that the covariates used in the analysis are time-varying or are constant.

This analysis was helpful in zeroing down the factors which are to be considered while selecting a particular model to define herding in Financial Markets. On comparing B Spline and Cox model it was found that since financial market data will not be left truncated and there will be chances of right censuring Cox model is a better choice. In the financial market, In case of a crash in the stock market or bank run the data collected to verify herding behaviour is not normally distributed and so it is not possible to assume the nature of the survival curve Cox Model seems to fit for the analysis. Even the proportionality assumption between hazard function and covariate can be tested after figuring out the covariate in the analysis.

References

1. Michael G. Hudgens, Ira M.Longini, Jr, M.Elizabeth Halloran, Kachit Choopanya, Suphak Vanichseni, Dwip Kitayaporn, Timothy D.Mastro, Philip A.Mock 2001- Estimating the Transmission probability of Human Immunodeficiency virus in Injecting Drug user in Thailand.Journal of Applied Statistics Vol 50, No1 pp 1-14
2. Michael G. Hudgens, Ira M.Longini, Suphak Vanichseni, Dale J Hu, Dwip Kitayaporn, Philip A.Mock, M.Elizabeth Halloran, Glen A Satten, Kachit Choopanya, Timothy D.Mastro - *Sub Type Specific Transmission probability of Human Immunodeficiency Virus Type 1 among Injecting drug user in Bangkok, Thailand.American Journal of Epidemiology, Vol 155, No 2, 2002*
3. Hamid Sabourian – the University of Toronto, Andreas Park-University of Cambridge- Herd Behavior in Efficient Financial Markets
4. Andrea Devenow, Ivo Welch – Rational Herding in Financial Economics.European Economic Review 40,1996.603-615
5. www.evidence-based-medicine.co.uk
6. Guang Guo –Event history analysis of left truncated data.Sociological Methodology Vol 23,(1993),217-243
7. http://en.wikipedia.org/wiki/Herding

