Nonparametric Hypothesis Testing for Isotonic Survival Models with Clustering

John D. Eustaquio
School of Statistics
University of the Philippines Diliman

A nonparametric test for clustering in survival data based on the bootstrap method is proposed. The survival model used considers the isotonic property of the covariates in the estimation via the backfitting algorithm. Assuming a model that incorporates the clustering effect into the piecewise proportional hazards model, simulation studies indicate that the procedure is correctly-sized and powerful in a reasonably wide range of scenarios. The test procedure for the presence of clustering over time is also robust to model misspecification.

Keywords: Bootstrap confidence interval; Survival Analysis; Clustered Data; Backfitting Algorithm; Generalized Additive Models; Nonparametric bootstrap.

1. Introduction

With the rise of big data and various data mining procedures, majority of service-oriented businesses are using these as ways for gathering insights and make-out patterns that would be beneficial for the company's growth and development. For these service-oriented businesses, the initial strategy for growth is to attract customers and then to convince them to maintain the customer-client relationship with their product or service in the long run

A lot of service-oriented sectors are extremely competitive, and the saturation of new customers happen very quickly. Thus, the recruitment of a customer (either a new entrant or a customer of a competing service provider) entails more cost compared to crafting promotional offers to retain present customers. Thus, it is vital for these businesses to develop ways of attracting customers and convincing them to become loyal customers. One method to maintain customers is to offer a lock-up period for the product that they availed (e.g., loyalty cards, telecommunication, credit cards, etc.). However, it is difficult to make the customers agree with the lock-in period without any significant incentives for them in turn. With this, there should be a balance with the value of the incentive and the lock-in period for certain customers. This optimization mechanism can be determined from the lifetime profitability of the customer.

This optimization problem is typically under the concern of Customer Relationship Management (CRM). One function of the CRM is to determine the churning of customers. Customer churn is the tendency of a customer to discontinue doing business with a service provider in a given period of time. This is crucial for the companies in the telecommunication, financial and other service and utility sectors. A task of the people in CRM is to know who among all the current customers has a relatively high probability of churning and to convince these customers to continue doing business with the company.

Churning can be brought about by several factors. Some of these factors can be considered as having an isotonic (monotonically increasing) effect on the customers churning rate. In the context of telecommunication companies, one example of this is the distance of the customer's area to the company's cellular site. This can be viewed as factor having an isotonic effect since as a customer's location goes farther away from the nearest cell site of a company, the weaker the communication signal the customer has, which would affect the churning of the customer. Another factor that has the isotonic is effect is the social contagion of a customer. As the number of people in contact with a customer who uses the same network provider decreases, there will be a corresponding increase in the customer's propensity to churn. Recent researches on social influence for customer acquisition and retention has shifted from giving support to the theory of social contagion to finding some explanations why it occurs (Iyengar et al., 2012). Some of the findings states that a customer is more likely to churn because that person becomes aware of a more attractive offer via information transfer through peers who have already churned. The client is also becoming aware that staying with the old provider will likely result in status disadvantages or peer pressure, or even realize the reduced cost that will be incurred as a result of the differential pricing of in-network and out-of-network calls due to the shift to other networks of the people in the customer's social network.

Aside from the factor with isotonic effect, some of these factors can be considered as clustering or grouping effects. These effects are those wherein members of the same cluster will more or less have similar characteristics that affect their churning rates. An example for this is the geo-demographic information of a customer. Braun and Schweidel (2011) studied the effects of geo-demographic clusters in the churning rate of customer. Table 1 shows the characteristics of each of the 5 clusters that they have studied.

Table 1. Characteristics of the Five Selected Demographic Groups

Cluster	Urbanicity	Income	Age	Has Kids	Housing	Employment	Educational Attainment
A	Urban	High	Middle	Yes	Rent	White Collar	College Grad
B	Town/Rural	High	Middle	Yes	Own	Management	College Grad
C	Suburban	Middle	Old	No	Own	Retired	Some College
D	Suburban	High	Middle	No	Own	Management	Post Grad
E	Urban	Middle	Young	Yes	Rent	White Collar	Some College

1.1 Rationale

A nonparametric survival model with clustering for predicting the survival time was postulated by Gauran and Barrios (2017). An additive model was used to characterize and understand the dynamics of customer churn:

$$\log\left[h_i^{(k)}(t)\right] = \lambda_k + \log\left[\lambda_i(t)\right] + \beta * f(x_i) + m * \varepsilon_i(t) \tag{1}$$

The logarithm of the response variable within the k^{th} cluster on time t, $\log\left[h_i^{(k)}(t)\right]$ is expressed as a linear function of λ_k, the cluster-specific random intercept; log [λ_i (t)], the baseline risk of the ith observation at time t; $f(x_i)$ the smooth function of the covariate of the ith observation; and $\varepsilon_i(t)$, the error term. It assumed in this model that the propensity to churn of the observation belonging in the same cluster are similar in behavior.

In the case of customer lifetime prediction, both the clustering factor and the isotonic effect of the covariate is very likely to occur. Thus, combining both models to account for both is deemed interesting. For this combined model, an assumption that the propensity to churn of each observation belonging in the same cluster are similar in behavior is needed.

It is also of great interest to determine when using this model if there is indeed a significant clustering effect. This study developed nonparametric procedures to verify the assumptions of the combined model. The inference procedures that are proposed are based on nonparametric bootstrap methods. Using nonparametric bootstrap methods allows to draw inferences with less demand for analytic formulas for the sampling distribution of the statistics entailed with the model and no need and strong assumption about the distributions (Mooney and Duval, 1993).

1.2 Objectives of the study

This study proposes a nonparametric isotonic regression for clustered survival models and to introduce a hypothesis testing procedure to test the significance of the clustering effect for the postulated model. This proposed modeling and inferential procedure of clustered time series data with some covariates having an isotonic effect is necessary because the scenarios wherein the model is applicable is normally occurring in consumer behavior study and customer lifetime prediction.

With the use of nonparametric bootstrap methods and hypothesis testing concepts more suitable for the bootstrap, the study aims:

1. To postulate and produce a nonparametric procedure for testing the presence of clustering effect in an isotonic clustered survival model; and
2. To determine the size and power of the proposed testing procedure for a desired level of significance.

1.3 Significance of the study

The market state of the mobile telecommunications sector has been rapidly growing in the early 2000s. Due to this rapid growth, in most countries, this sector has almost attained the point of saturation in terms of the slow growth of mobile subscribers. Because of this, telecommunication companies changed their plans and strategies. Customer retention became the focus rather than customer acquisition since acquisition of new customers entails higher cost than keeping the existing customers loyal when the derived income from the customers is taken into consideration.

From the managerial point of view, the social network of a customer can provide pertinent information on the subject of churn prediction (Haenlein, 2013). Even a small improvement in the accuracy of churn prediction models can lead to substantial financial gains for companies (Neslin, et al., 2006). This is applicable in the telecommunication industry since customer churn has been shown to be a large cost item while the information regarding the social relationship of a customer can be easily obtained through the details of the customer's call records.

It is also of interest to know if there are some factors that affects the churning among customer sub-groups or clusters. Companies can incur lower cost by giving promotions to large groups of customers instead of targeting specific individuals if they know that a clustering effect is present. Thus, it is necessary to create a way to make inference regarding the existence of clustering effect.

The testing procedures developed in this can be used to closely examine various clustering factors that may be present in a survival data. There are no assumptions made on the sampling distributions of the statistics used in hypothesis testing since the bootstrap procedure is used to approximate these sampling distributions. Because of this, the execution of the said procedures is relatively simple and straightforward despite being computer-intensive with the incorporation of bootstrap methods for inference. The proposed testing procedures yielded reliable results provided that the bootstrap is consistent.

2. Related Literature

2.1 Inferences for Cox's regression models

The use of the Cox proportional hazard model is popular in the modelling events in the context of survival analysis. This is because the Cox's regression model can quantify the relationship between failure times and a set of covariates. The equation for the Cox proportional hazard model given by Cox (1972) is as follows:

$$\lambda(t \mid Z) = \lambda_0(t) exp(\beta_0^t Z)$$

where λ_0 is an unknown baseline hazard function, Z is a p-dimensional covariate and β_0 is a vector of regression coefficients.

2.2 Isotonic regression models for survival data

In medical studies, it is of interest to model clinical events in relation to disease progression using various continuous variables. In most cases, it is sensible to make the assumption that the odds of disease progression follow a monotone function of the continuous variable used in modelling. Also in most cases, the continuous variable is an indicator which is recorded for a period of time which is longitudinal in nature. Ancukiewicz (2003) proposed two methods for modelling the relationship of the hazard and the covariate. The first one is to assume that the hazard is constant throughout time while the second method allows the hazard to be an arbitrary function of time.

The second method proposed by Ancukiewicz is modified by Ma (2010) where the isotonic property of the covariate in relation to the hazard is more flexible. The result of the isotonic model will be step function were some steps of Y going up as X increases will be relatively small and hence insignificant. Thus, some of the adjacent step can be combined if they are not significantly different from each other.

2.3. Hypothesis testing procedures via bootstrap

When the asymptotic distribution of a test statistic is difficult or almost impossible to derive analytically, bootstrap based inferences are commonly used. These cases wherein derivation of the distribution of the test statistic is difficult may arise from unobserved heteroskedasticity or the presence of nuisance parameters that affects the test statistic.

Guarte and Barrios (2013) proposed a way to conduct nonparametric hypothesis for spatio-temporal models. The procedure makes use of the bootstrap confidence interval in order to test for the spatial and/or the temporal effect in the model since it assumes that each component in the model has an additive effect. This assumption makes the hypothesis testing procedure more efficient to use since the effect of each components can be separated from each other easily.

3. METHODOLOGY

3.1. Nonparametric model for clustered customer survival data with isotonic covariate

In the literature of survival analysis, it is said that when the data at hand is comprised in forms of clusters, the existing parametric models are not optimal. Thus, modeling via nonparametric methods is proposed by Gauran and Barrios (2017) to provide a better characterization for the churning of the observations in a clustered survival data.

Given n clusters with m elements observed for T time points, the postulated isotonic model:

$$\log\left[h_i^{(k)}(t)\right] = \lambda_k + log\left[\lambda_i(t)\right] + g(x_i)$$

where $i = 1,2,\ldots, m$,

n is the total number of clusters

$h_i^{(k)}(t)$ is the i^{th} value of the response variable within the k^{th} cluster on time t

λ_k is the cluster-specific random intercept

λ_i (t) is the baseline risk of the ith observation at time t

x_i is the value of the covariate of the i^{th} observation within the k^{th} cluster

$g(x_i)$ is the isotonic function of x_i equal to g_s for $x_{s-1} \leq x_i < x_s$ such that s is defined at specific ranges of the covariate.

Backfitting algorithm is employed in the estimation of the terms in the model since the postulated model has additive components. Since the interest in the long run is to make inferences on the clustering effect, this is the first term that would be estimated in the backfitting process. Afterwards, the components were estimated sequentially from the most important term up to the least important term in the model. The backfitting estimation procedure is iterated until there are no more significant information for regarding the clustering effect left on the residuals which indicates convergence.

3.2 Estimating the clustered isotonic survival model

A procedure is proposed by Eustaquio and Barrios (2016) to test the clustering effect in a nonparametric survival model using bootstrap confidence intervals. This procedure was combined with the pooled adjacent violators algorithm (PAVA) for estimating isotonic models. The clustered isotonic survival model is estimated as follows:

1. The function of the covariates, $g(x_i)$ and the baseline risk, $log[\lambda_i(t)]$ are ignored first and the cluster specific term, λ_k is estimated. Since λ_k is a random effect, a mixed effects model will be used to estimate $\widehat{\lambda_k}$.

2. The partial residual $\widehat{e_i}$ is computed as $\left(\log\left[h_i^{(k)}(t)\right]-\widehat{\lambda_k}\right)$. At this point, the partial residual contains information on $g(x_i)$ and $log[\lambda_i(t)]$.

3. Ignore $g(x_i)$ further, the partial residual $\widehat{e_i}$ will be used to estimate $log[\lambda_i(t)]$. The residual is computed again as $\left(\log\left[h_i^{(k)}(t)\right]-\widehat{\lambda_k}-\widehat{log[\lambda_1(t)]}\right)$. This residual is a function of $g(x_i)$ alone.

4. Again, the residual will be used to estimate $g(x_i)$. The function of the covariate is estimated using PAVA

   a) If $g(x_i)$ is in non-decreasing order the $g^*(x_i) = g(x_i)$.
   b) If the is a violator such that $g(x_i) > g(x_{i+1})$ for some x_i. Replace these two values by their weighted average: $Av\{g(x_i), g(x_{i+1})\}$.
   c) Afterwards, x_i, x_{i+1} form a block called *level set* (LS). If the new set of K-1 values is isotonic, then $g^*(x_i) = g^*(x_{i+1}) = Av\{g(x_i), g(x_{i+1})\}$ for the violator and $g^*(x_i) = g(x_i)$ for all other observations
   d) If the set is not isotonic, repeat the procedure using the new set of values until the entire set is isotonic

5. Finally, the residual is computed as $\left(\log\left[h_i^{(k)}(t)\right]-\widehat{\lambda_k}-\widehat{log[\lambda_1(t)}-\widehat{g(x_l]}\right)$. The estimate for component to be estimated in the next step is left out for the computation of the residual to ensure that the residual will contain information on that term.

6. Iterate the procedure above until convergence.

3.3 Hypothesis testing procedure for the presence of cluster effect

Given the survival data in each cluster, we test the following hypotheses:
Ho: All cluster have the same churning effect: $\lambda_1 = \lambda_2 = \cdots = \lambda_k$
Ha: At least one cluster differs in its churning effect: $\lambda_i \neq \lambda_j$
for at least one $i \neq j$

Based on the chosen bootstrap method for regression analysis, the following procedure for testing the hypotheses proposed by Eustaquio and Barrios (2016) is applied.

Algorithm:

1. Estimate the clustered survival model assuming that all the observation are from the same cluster.

2. Estimate λ_0, the cluster-specific random variable assuming all the observation belong in the same cluster produced by the estimated model.

3. Generate r bootstrap resamples of size n from the data.

4. Estimate the model for each of the resamples.

5. Compile each λ_k's estimated by each resamples.
6. Sort each group of λ_k (i.e λ_1, λ_2, λ_3) in either ascending or descending order.
7. Find the appropriate percentile to construct a (1-α) % confidence interval.
8. Reject the null hypothesis that there is no clustering effect at α% level of significance if at least one of the intervals fail to contain λ_0.

The procedure is deemed appropriate for testing the clustering effect across all clusters for any clustered survival data. However, the comparison of the clustering effect is relies solely on one parameter, λ_k in the model.

3.4 Simulation studies

A simulation study is designed to evaluate the power and size of the proposed testing procedure for clustering effect of the clustered survival model. For each simulation setting, the hypothesis testing procedure is repeated 200 times to determine the reliability of the proposed test. Table 2 shows the simulation boundaries employed in the study.

Table 2. Boundaries of Simulation Study

1	Number of Clusters	Small – 3 Large – 10
2	Distribution of the error term	$\varepsilon \sim N(0,1)$ $\varepsilon \sim N(0,10)$
3	Misspecification m in the model	$m = 1$, $m = 5$
4	Distribution of $\delta_i\,(t)$	$\delta_i\,(t) \sim Exp(1)$
5	Distribution of x_i	$x_i \sim$ Uniform (-1,1)
6	Functional form $g(x_i)$	Linear – aX + b Sigmoidal – aΦ (2.5X) Exponential – ae^X where a and b are constants
7	Distribution of λ_k	(see table 3)

Based on Gauran and Barrios (2017), the overall behaviour of the model cannot be understood without knowing the specification of the baseline hazard function. Thus, $log[\lambda_i\ (t)]$ is assumed to $\gamma e^{-\gamma t} + \delta_i\,(t)$ where $\gamma = 0.5$. The scale parameter of the exponential distribution is the constant churn rate. Even if the number of observations (cluster size) in each cluster is constant, the number of clusters is allowed to vary (small=3, large=10). These simulation settings aims to address the fact that adding clusters not individual observations leads to an increase in efficiency (Arceneaux and Nickerson, 2009).

Poisson distribution is used as the distribution of the clustering effect with varying means for each clusters as indicated in Table 3. These settings were considered to account the various cases of the clustering effect, from those with very close means up to those with values that are very far apart. The settings I and J were considered to be able measure the size of the proposed test.

Table 3. Mean of the Poisson Distributed Cluster-Specific Component

Scenario		No. of Clusters	Cluster no.									
			1	2	3	4	5	6	7	8	9	10
A	With Clustering Effect	Small k=3	5	6	7	--------------------------------------						
B			5	7	9	--------------------------------------						
C			5	9	13	--------------------------------------						
D			5	5	10	--------------------------------------						
E		Large k=10	5	7	9	11	13	15	17	19	21	23
F			5	5	5	5	5	5	6	6	6	6
G			5	5	5	5	5	5	10	10	10	10
H			5	5	5	7	7	7	7	9	9	9
I	Without Clustering Effect	Small	5	5	5	--------------------------------------						
J		Large	5	5	5	5	5	5	5	5	5	5

There are two distributions that are examined for the error term. The first one portrays the scenario where there is minimal variation while the other takes into account the possibility of high variability or heterogeneity among the observations. The constant m which is multiplied to the error term, is included to explore the effect on the hypothesis testing procedure of the model if there were a misspecification error.

4. Results and Discussion

For each of the simulation boundaries stated above, the results were presented by the size of the cluster (n=10 and n=100 for small and large cluster size respectively). The value m, an index of the extent of misspecification error and the variance of the error term are also included in the disaggregation of the results.

4.1 Power of the test for the presence of clustering effect

The power of the proposed test is computed for the scenarios with small or large number of clusters where at least one of the means of the cluster-specific term is different from the other. For the computation of the power, the test is conducted for 200 times in every settings where the null hypothesis is indeed false and the proportion of that resulted to the rejection of the null hypothesis is deemed the power of the test for each setting.

4.1.1 Small number of clusters

Four major scenarios were considered in measuring the power of the test for small number of clusters (k=3). Ranging from those with the distribution of the cluster-specific that have really close means up to those with one cluster having the mean of the cluster-specific term very far away from the mean of the other two clusters.

Mean increments by 1 for each succeeding cluster

Table 4 shows that the power of the test is about 0.90 for cases where there is no misspecification and the variance of the error term is small. In cases where there is misspecification and the variance of the error term is high, the power of the proposed test in these scenarios is near to 0.85.It also indicates that the test is relatively more powerful in most cases that the sample size is large compare to those with smaller sample size.

Table 4. Power of the Test when the Mean Increments by 1 for each Succeeding Cluster

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_{ij})$	*Linear*	**misspecification (m)**	$m = 1$	0.945	0.975	0.875	0.890
			$m = 5$	0.905	0.955	0.835	0.830
	Sigmoidal		$m = 1$	0.900	0.920	0.910	0.920
			$m = 5$	0.855	0.875	0.830	0.860
	Exponential		$m = 1$	0.960	0.990	0.915	0.920
			$m = 5$	0.945	0.965	0.895	0.910

Mean Increments by 2 for Each Succeeding Cluster

When the means of the cluster specific term are moderately far from each other, the power of the test are mostly above 0.95. The power of the test for this cases is not much affected by misspecification when the variance of the error term is low, but in some cases where there is high variability and misspecification is present, the power of the test dropped significantly but still not lower than 0.85.

Table 5. Power of the Test when the Mean Increments by 2 for each Succeeding Cluster

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_{ij})$	*Linear*	**misspecification (m)**	$m = 1$	0.965	0.985	0.905	0.910
			$m = 5$	0.930	0.960	0.850	0.850
	Sigmoidal		$m = 1$	0.930	0.945	0.940	0.955
			$m = 5$	0.870	0.895	0.850	0.875
	Exponential		$m = 1$	0.970	1.000	0.950	0.935
			$m = 5$	0.965	0.980	0.935	0.955

Mean increments by 4 for each succeeding cluster

The proposed testing procedure that yielded very high power which is very close to one in almost all cases when the means of the cluster-specific term are quite far from each other. The test resulted in a higher power even if misspecification occurred and the variability of the error term is large.

Table 6. Power of the Test when the Mean Increments by 4 for Each Succeeding Cluster

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_{ij})$	*Linear*	**misspecification (m)**	$m = 1$	0.985	1.000	0.970	0.995
			$m = 5$	0.965	0.985	0.950	0.975
	Sigmoidal		$m = 1$	0.960	0.975	0.945	1.000
			$m = 5$	0.940	0.950	0.950	0.935
	Exponential		$m = 1$	0.990	1.000	0.975	0.995
			$m = 5$	0.985	1.000	0.960	0.980

Two clusters have the same effect and the other is quite different

The power of the test for most of the cases where the means of two out of the three cluster-specific term are the same and the other one is quite far from the other two is at least 0.9. Still consistent with the literature, the lowest power of the test for this scenario is attained when there is misspecification and the variability is high.

Table 7. Power of the Test when Two Clusters have the Same Effect and the Other is Quite Different

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_{ij})$	*Linear*	**misspecification (m)**	$m = 1$	0.985	1.000	0.970	0.995
			$m = 5$	0.965	0.985	0.950	0.975
	Sigmoidal		$m = 1$	0.960	0.975	0.945	1.000
			$m = 5$	0.940	0.950	0.950	0.935
	Exponential		$m = 1$	0.990	1.000	0.975	0.995
			$m = 5$	0.985	1.000	0.960	0.980

4.1.2 Large number of clusters

For the scenario with fairly large number of clusters (k=10), 6 scenarios were considered in determining the power of the test. Ranging from those with the distribution of the cluster-specific term that have really close means, to those with one cluster that has the mean of the cluster-specific term very far away from the mean of the other 9 clusters.

Mean increments by 2 for each succeeding cluster

When the mean of the cluster-specific terms is increased by 2 for each of the succeeding clusters, the power of the test is very near to 1 in most cases. Even when there is misspecification, the power of the test is around 0.90. Thus, the test is still powerful even if there is misspecification and the variability is high.

Table 8. Power of the Test when the Mean Increments by 2 for Each Succeeding Cluster

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_i)$	*Linear*	misspecification (m)	$m = 1$	0.965	0.985	0.905	0.910
			$m = 5$	0.930	0.960	0.850	0.850
	Sigmoidal		$m = 1$	0.930	0.945	0.940	0.955
			$m = 5$	0.870	0.895	0.850	0.875
	Exponential		$m = 1$	0.970	1.000	0.950	0.935
			$m = 5$	0.965	0.980	0.935	0.955

Clusters have two inherent grouping effect that are close to each other

In reference to table 9, the power of test is low in most cases when the cluster means are almost the same for all clusters. Consistent with the results of the other scenarios, the power increases when there is no misspecification and the variability is low. But even for the cases where there is no misspecification, the test is still not ideal for this detecting the presence of clustering effect in this scenario.

Table 9. Power of the Test when the Clusters have Two Inherent Grouping Effect that are Close to Each Other

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_i)$	*Linear*	misspecification (m)	$m = 1$	0.575	0.610	0.445	0.515
			$m = 5$	0.320	0.535	0.300	0.315
	Sigmoidal		$m = 1$	0.930	0.945	0.940	0.955
			$m = 5$	0.870	0.895	0.850	0.875
	Exponential		$m = 1$	0.575	0.615	0.455	0.520
			$m = 5$	0.340	0.540	0.320	0.345

Clusters have two inherent grouping effect that are far from each other

As the cluster-specific terms' mean for large number cluster groups into two values but are very far from each other, the power of the test is close to 0.85 in most cases except for the scenario where there is misspecification and large variability. In those cases where there is misspecification and the variability is high, the power of the test decreased but is still near to 0.8.

Table 10. Power of the Test when the Clusters have Two Inherent Grouping Effect that are Far from Each Other

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_i)$	*Linear*	misspecification (m)	$m = 1$	0.855	0.935	0.810	0.815
			$m = 5$	0.815	0.825	0.760	0.780
	Sigmoidal		$m = 1$	0.870	0.945	0.830	0.835
			$m = 5$	0.825	0.855	0.755	0.810
	Exponential		$m = 1$	0.850	0.905	0.800	0.795
			$m = 5$	0.800	0.790	0.740	0.790

Clusters have three inherent grouping effect that are far from each other

The proposed test had the power of about 0.6 for most of the cases where the means of the cluster-specific term are almost the same for most of the cluster while the other clusters have equal means, though their mean is quite far from the other group of clusters. Still consistent with the literature, the lowest power of the test for this scenario is attained when there is misspecification and the variability is high. In this case, it can also be seen that decreasing the level of significance led to a significant decrease to power of the test.

Table 11. Power of the Test when the Clusters have Three Inherent Grouping Effect that are Far from Each Other

Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_i)$	*Linear*	misspecification (m)	$m = 1$	0.675	0.760	0.560	0.595
			$m = 5$	0.525	0.655	0.380	0.45
	Sigmoidal		$m = 1$	0.645	0.730	0.530	0.565
			$m = 5$	0.510	0.625	0.360	0.370
	Exponential		$m = 1$	0.730	0.760	0.570	0.605
			$m = 5$	0.555	0.645	0.435	0.490

4.2 Size of the test for the presence of clustering effect

The proposed test is also assessed for its size by using two scenarios, one with small and the other one with large number of clusters where the cluster-specific term all comes from the Poisson distribution with the same means. To check whether the test is correctly-sized, the test is conducted 200 times in each of the settings where the null hypothesis is true, there is no clustering effect. If more than 100*α of the results rejected the null hypothesis, the test is considered incorrectly-sized for that setting.

4.2.1 Small number of clusters

In the case where there really is no clustering effect for small number of clusters (k=3), the test is correctly sized all of the cases when there is a 10% level of significance even if there is misspecification and the data is highly heterogenous. The size of the test for this scenario is still correct even if the cluster size is small (n=10).

For small number of clusters, at 5% level of significance, the test is still correctly sized for most of the cases where the cluster –specific term comes from the same distribution for all of the three clusters. The test became incorrectly sized when the cluster size is small. Also, the test is incorrectly-sized when the functional form of the covariate is exponential or sigmoidal and there is misspecification. This could be attributed to the fact the misspecification has the chance to increase the heterogeneity of the data on hand.

4.2.2 Large number of clusters

The proposed test is correctly-sized for most of the cases where there are large number of clusters and the distribution of the cluster-specific term is the same across all ten clusters when 10% level of significance is employed. Still consistent with the previous results for the size when there is a small number of clusters, the test became incorrectly-sized when there is misspecification in a highly heterogenous data. In this case, it can also be seen that the test became incorrectly-sized from the cases mentioned earlier when the cluster size is large that might be attributed to reason stated in the previous results.

The test is correctly sized when the functional form of the covariates is linear. For the case where the functional form of the covariates is exponential or sigmoidal, the test is incorrectly sized when the cluster size is large at 5% level of significance. Also consistent with the results of the other scenarios, the test turns out to be incorrectly-sized when there is misspecification and/or the cluster size is large for the cases.

As for the case where the mean of the cluster-specific term comes from the same distribution for large number of clusters, the test is correctly sized in most of the cases at 1% level of significance. In those cases, where there is misspecification and the cluster size is large, the test also became incorrectly sized which is consistent with the other scenarios simulated previously.

5. Conclusions

The proposed procedures are able to properly categorize the actual behavior (constant and not constant clustering effect) of the clustering effect. Simulation studies show that the testing procedure has a high power and is correctly sized in various scenarios of clustered survival data. The power of the test is higher (close to 1 or equal to 1) for those scenarios with distant alternative values, both for small and large number of clusters for those with covariates with linear effect.

For cases where the clustering effect is quite near for each cluster even though at least one cluster has a quite different effect, the testing procedure still resulted with a high power.

In survival data with a large number of clusters, even with highly heterogenous data and misspecification is present, the test is still powerful even if the covariates have a nonlinear isotonic effect. To be able to conduct the testing procedure, the needed data should have the following information 1) the survival time of each observation, 2) the covariates involved in the model, 3) and the classification to which cluster each observation belong.

REFERENCES

ANCUKIEWICZ M., FINKELSTEIN D.M. and SCHOENFELD D.A., 2003 Modeling the relationship between continuous covariates and clinical events using isotonic regression. *Statistics in Medicine* 22, 3151-3159.

ARCENEAUX, K. and NICKERSON, D., 2009, Modeling Certainty with Clustered Data: A comparison of Methods. *Political Analysis*, 17, 177-190.

BRAUN, M., AND SCHWEIDEL, D. A., 2011, Modeling Customer Lifetimes with Multiple Causes of Churn, *Marketing Science*, 30:5, 881–902.

COX, D.R., 1972, Regression model and life tables (with discussion), *Journal of Royal Statistical Society B*, 34,187–220.

EUSTAQUIO, J. D. and BARRIOS, E.B., 2017, *Nonparametric Hypothesis Testing in Clustered Survival Model. Forthcoming in Communications in Statistics-Simulation and Computation*, DOI: 10.1080/03610918.2016.1241407. [Taylor & Francis Online]

GAURAN, I. M. and BARRIOS, E.B., 2017, Nonparametric Modeling of Clustered Customer Survival Data,*Communications in Statistics-Simulation and Computation*, 46(1):603-618.

GUARTE, J.M. and BARRIOS, E.B., 2013, Nonparametric Hypothesis Testing in a Spatial-Temporal Model: A Simulation Study, *Communications in Statistics - Simulation and Computation*, 42:1, 153-170.

HAENLEIN, M., 2013, Social interactions in customer churn decisions: The impact of relationship directionality, *International Journal of Research in Marketing*, 30:3, 236 – 248.

IYENGAR, R., VAN DEN BULTE, C., and CHOI, J., 2012, Distinguishing among mechanisms of social contagion in new product adoption: Framework and illustration. Working paper: The Wharton School.

MOONEY, C.Z. AND DUVAL, R.D., 1993, *Bootstrapping: A Nonparametric Approach to Statistical Inference*. CA: SAGE Publications, Inc.

NESLIN, S. A., GUPTA, S., KAMAKURA, W., LU, J., & MASON, C. H., 2006, Defection detection: Measuring and understanding the predictive accuracy of customer churn models, *Journal of Marketing Research*, 43:2, 204–211.

MA Y. (2010) Flexible Isotonic Regression in Survival Data Analysis. PhD Dissertation. The George Washington University.

Appendix

Table 12. Size of the test for the settings when there are small number of clusters (k=3)

Level of Significance				$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_{ij})$	*Linear*	**misspecification (m)**	$m = 1$	3	7	6	6	0	3	1	5	0	0	0	0
			$m = 5$	6	6	9	6	3	1	4	2	0	1	1	1
	Sigmoidal		$m = 1$	6	3	5	9	0	2	3	5	0	0	0	1
			$m = 5$	6	7	9	10	4	2	5	4	1	2	1	3
	Exponential		$m = 1$	7	9	9	10	5	4	5	6	0	3	1	3
			$m = 5$	9	7	5	10	3	6	1	7	0	3	1	4

Table 13. Size of the test for the settings when the number of clusters is large (k=10)

Level of Significance				$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
Distribution of Error Term				$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$		$\varepsilon \sim N(0,1)$		$\varepsilon \sim N(0,10)$	
Sample Size				$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$	$n = 10$	$n = 100$
Functional Form of $g(x_{ij})$	*Linear*	**misspecification (m)**	$m = 1$	9	5	9	8	5	3	5	5	0	0	0	0
			$m = 5$	8	6	8	9	4	4	3	5	0	1	1	3
	Sigmoidal		$m = 1$	8	9	6	9	2	4	3	3	0	0	0	2
			$m = 5$	7	8	7	10	2	5	4	8	0	3	1	5
	Exponential		$m = 1$	7	7	10	10	2	5	5	6	1	2	1	4
			$m = 5$	6	9	7	15	4	7	5	10	1	5	1	8